Establishing site-specific recombination mediated cassette exchange in *Pichia pastoris* and development of an optimized process for the production of mouse Tmprss2

Von der Fakultät für Lebenswissenschaften
der Technischen Universität Carolo-Wilhelmina zu Braunschweig
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)
genehmigte
D i s s e r t a t i o n

von Christian In-Shu Schinkowski
aus Braunschweig

1. Referent:	Professor Dr. Ralf-Rainer Mendel
2. Referent:	Professor Dr. Wulf Blankenfeldt
Eingereicht am:	20.02.2017
Mündliche Prüfung (Disputation) am:	05.04.2017
Druckjahr	2017

Vorveröffentlichungen der Dissertation

Teilergebnisse aus dieser Arbeit wurden mit Genehmigung der Fakultät für Lebenswissenschaften, vertreten durch den Mentor der Arbeit, in folgenden Beiträgen vorab veröffentlicht:

Tagungsbeiträge

Bock T., Schinkowski C., Schürig M. (all authors contributed equally to this presentation)
"From 'Gene' to Application"
HZI Grad School, 6th Annual Retreat, Goslar-Hahnenklee, Germany (2015)

Schinkowski C., Schürig M., Stricker R., Krauße J. & van den Heuvel J.
"Development of New Pichia pastoris RMCE Master Cell Lines to Produce Difficult to Express Proteins"
HZI, Progress Seminar, Braunschweig, Germany (2014)

Posterbeiträge

Schinkowski C., Meyer S. & van den Heuvel J.
"Development of New Pichia pastoris RMCE Master Cell Lines to Produce Difficult to Express Proteins"
HZI Grad School, 8th International PhD Symposium, Braunschweig, Germany (2015)

Schinkowski C., Schürig M., Stricker R., Krauße J., Schughart K., & van den Heuvel J.
"Expression of Hemagglutinins and Tmprss2 for Novel Strategies Against Influenza Infection"
IGBMC Symposium New strategies for macromolecular complexes analysis, Strasbourg, France (2014)

Schinkowski C., Schürig M., Stricker R., Krauße J., Schughart K., & van den Heuvel J.
"Expression of Hemagglutinins and Tmprss2 for Structural Biology to Develop Novel Strategies Against Influenza Infection"
HZI Grad School, 7th International PhD Symposium, Braunschweig, Germany (2014)

Schinkowski C., Meyer S., & van den Heuvel J.
"Development of New Pichia pastoris RMCE Master Cell Lines to Produce Difficult to Express Proteins"
HZI Grad School, 5th Annual Retreat, Goslar-Hahnenklee, Germany (2014)
(3rd Poster Award)

Schinkowski C., Meyer S., & van den Heuvel J.
"Development of New Pichia pastoris RMCE Master Cell Lines to Produce Difficult to Express Proteins"
HZI Grad School, 6th International PhD Symposium, Braunschweig, Germany (2013)
(3rd Poster Award)

Table of Contents

Abbreviations

aa	Amino acids
(m/p)Ab	(monoclonal/polyclonal) Antibody
ACT1	Beta-actin gene
AF	Anti-Foam
Amp	Ampicillin
AOX	Alcohol oxidase
AP	Alkaline phosphatase
ARG4	Argininosuccinate lyase
ATG	Translation start codon
A_{260} / A_{280}	Absorbance at 260 nm / 280 nm
bp	Base pair
BCIP	5-Bromo-4-chloro-3-indilyl phosphate
BEVS	Baculovirus expression vector system
Bluo-Gal	5-Bromo-3-indolyl-β-D-galacto pyranoside
BMD(Y)	Buffered minimal dextrose medium (with yeast extract)
BMG(Y)	Buffered minimal glycerol medium (with yeast extract)
BMM(Y)	Buffered minimal methanol medium (with yeast extract)
BSA	Bovine serum albumine
CXCR4	C-X-C chemokine receptor 4
CD4	Cluster of differentiation 4
CDS	Coding sequence
CHO	Chinese Hamster Ovary
Cm	Chloramphenicol
C_T	Cycle threshold
CMV	Cytomegalovirus
CUB	Cls/Clr, urchin embryonic growth factor bone morphogenetic protein-1
CV	Column volume
ddH2O	H_2O bidest
DESC	Differentially expressed in squamous cell carcinoma
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotides (A, T, C, G)
DTT	Dithiothreitol
ε	Molar extinction coefficient [$L \cdot g^{-1} \cdot cm^{-1}$]
EBNA1	Epstein-Barr Nuclear antigen 1
EBV	Epstein-Barr virus
E. coli	*Escherichia coli*
ECD	Extra cellular domain
EDTA	Ethylene diamine tetra acetic acid
ER	Endoplasmic reticulum
eGFP	Enhanced green fluorescent protein
EM7	Synthetic promoter derived from T7 promoter
ESI	Electrospray ionization
EtOH	Ethanol
Fab	Fragment antigen binding
Fc	Fragment crystallizable region of an antibody
FCS	Fetal calf serum
Flp	Flp recombinase / Flippase
FRT	Flp recombinase targeting
FSC	Forward Scatter
GAP	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
Gm	Gentamicin
GOI	Gene of interest
HA (HA0 / HA1, HA2)	Hemagglutinin (unprocessed / subdomains 1 and 2)
HAT	human airway trypsin-like protease

HEK	Human Embryonic Kidney
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
Hi5	Trichoplusia ni BTI 5B1-4 insect cell line
His4 / HIS4	Histidine biosynthesis trifunctional protein
HIV	Human immunodeficiency virus
HMW	High molecular weight
H1 / H3 / H7	Hemagglutinin 1 / 3 / 7
H1N1 / H3N2 / H7N7	Hemagglutinin 1/3/7 neuraminidase 1/2/7
hPGK	Human PGK1
hpi	Hours post induction (with methanol)
hpt	Hours post transfection
HR	Homologous recombination
HRP	Horseradish peroxidase
HSA	Human serum albumin
HZI	Helmholtz-Centre for Infection Research, Braunschweig, Germany
IAV	Influenza A virus
IPTG	Isopropyl-β-D-thiogalactopyranoside
Kan	Kanamycin
kbp	Kilo base pair (1000 bp)
kDa	Kilodalton
L	Litre
LB	Lysogeny broth
LDLA	Low density lipoprotein receptor class A
LMW	Low molecular weight
M	Molarity [mol/L] // For gel images: Molecular weight standard
MALDI	Matrix assisted laser desorption/ionisation
MAM	meprin/A5 antigen/receptor protein phosphatase mu
MCL	Master cell line
MCS	Multiple cloning site
MD(H)	Minimal dextrose medium (supplemented with histidine)
MeOH	Methanol
MF-α ss	alpha mating factor pre pro secretion signal
mHost-XS	multi host expression system
MQ-H_2O	MilliQ H_2O
MS	Mass spectrometry
MSLP	Mosaic serine protease large-form
mut^S / mut^+ / mut^-	methanol utilization slow / plus / negative
MW	Molecular weight
MWCO	Molecular weight cut-off
NaCl	Sodium chloride
NBT	Nitro blue tetrazolium chloride
NHEJ	non-homologous-endjoining
Ni-NTA	Complex of nickel ions and nitriloacetic acid
NPV	Nucleopolyhedrovirus
NRT	No reverse transcriptase control (RT-PCR)
NTC	No template control (NTC)
OD_{595} / OD_{600}	Optical density (at 595 nm / 600 nm)
OpIE	*Orgyia pseudotsugata* immediate early
ORF	Open reading frame
ori	Origin of replication
oriP	Origin of replication (EBV)
P. pastoris	*Pichia pastoris (Komagataella phaffii)*
PAGE	Polyacrylamide gel electrophoresis
PARS	Pichia autonomous replicating sequence
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
PEP4	Carboxypeptidase Y-deficient protein 4
PES	Polyether sulfone
PGK	Phosphoglycerate kinase

PhD	Scientific doctoral degree (lit. Doctor of Philosophy)
PI	Propidium iodide
pI	Isoelectric point
polyA	Polyadenylation signal
PVDF	Polyvinylidene difluorid
RFP	Red fluorescent protein
(m)RNA	Messenger ribonucleic acid
RMCE	Recombinase mediated cassette exchange
rpm	Revolutions per minute
RT-PCR	(Quantitative) reverse transcription PCR
SARS	Severe acute respiratory syndrome
SBDD	Structure-based drug design
Sc	*Saccharomyces cerevisiae*
ScFv	Single chain variable fragment of an antibody
ScARG4	ARG4 gene from *Saccharomyces cerevisiae*
SCP	Single-cell protein
SDS	Sodium dodecyl sulphate
SEA	Sea urchin sperm protein/enteropeptidase/agrin
SF	*Spodoptera frugiperda*
SOC	Super optimal broth for catabolite repression
SPD	Serine protease domain
SRCR	Scavenger receptor cysteine-rich domain
SSC	Sideward scatter
TAE	Tris-acetate-EDTA
TBS-T	Tris buffered saline with Tween-20
TCA	Trichloroacetic acid
TE	Tris-EDTA
TEF1	Translation elongation factor EF-1 alpha
TEMED	N,N,N´,N´-Tetramethylethylendiamin
Tet	Tetracyclin
TEV (protease)	Tobacco Etch Virus (nuclear-inclusion-a endopeptidase)
Tm	Melting temperature
TM	Transmembrane domain
Tmprss	Transmembrane protease serine
Tn7	Bacterial transposon Tn7
TOF	Time of flight
Tris	2-Amino-2-hydroxymethyl-propane-1,3-diol
TTSP	Type two transmembrane protease
U	Catalytical units
UPR	Unfolded protein response
URG	Genomic upstream region of GAP
UV	Ultraviolet
VEGF	Vascular endothelial growth factor
v/v	Volume fraction (volume per volume)
w/v	Mass concentration (weight per volume)
WT	Wildtype
x g	Gravitational acceleration
y-eGFP	Yeast-enhanced green fluorescent protein
YFP	Yellow fluorescent protein
YNB	Yeast nitrogen base
YPD(S)	Yeast extract peptone dextrose medium (with 1 M sorbitol)
Zeo	Zeocin

Zusammenfassung

Die Durchführung strukturbiologischer und biochemischer Analysen erfordert große Mengen an hochreinen Zielproteinen. Dabei steigt die Nachfrage nach komplexen, schwer zu exprimierenden Proteinen stetig, deren Produktion jedoch einen entscheidenden Engpass darstellt. Die Optimierung von Expressionssystemen und Produktionsprozessen ist daher ein wichtiges Unterfangen. Die methylotrophe Hefe *Komagataella phaffii* (*Pichia pastoris*) ist ein verbreitetes Expressionssystem zur Herstellung rekombinanter Proteine. Sie vereint die kostengünstige und einfache Handhabung von Mikroorganismen mit der Fähigkeit eine Vielzahl der posttranslationalen Modifikationen höherer Eukaryoten durchzuführen. Dadurch stellt *P. pastoris* eine interessante Alternative zu kostenintensiven Zellkulturen dar. Jedoch ergeben sich bei der erforderlichen stabilen Transformation von *P. pastoris* über homologe Rekombination einige entscheidende Nachteile. Diese umfassen eine begrenzte Anzahl etablierter genomischer Loci zur Integration von Zielgenen und eine stark variierende Expressionsleistung transformierter Klone, wodurch in der Regel ein langwieriges *Screening* nach Hochproduzenten („Jackpot-Klonen“) notwendig ist.

Der erste Teil dieser Arbeit behandelt die Etablierung eines Rekombinase vermittelten Kassettenaustauschsystems (*recombinase mediated cassette exchange*, RMCE) zur Optimierung des Expressionssytems *P. pastoris*. Die spezifische Integration von Zielgenen in eine definiert austauschbare genomische Kassette einer *P. pastoris* RMCE *Master*-Zelllinie durch das Enzym Flp Rekombinase stellt eine vielversprechende Option dar, das *Screening* nach Jackpot-Klonen durch vorhersagbare Produktionsraten zu verhindern und bisher ungenutzte genomische Loci zur stabilen Proteinexpression zu erschließen. In dieser Arbeit wird die Herstellung eines RMCE Modell-Systems in *P. pastoris* gezeigt. Ein beobachteter kritischer Einfluss einer Erkennungssequenz der Flp Rekombinase (FRT F_3) auf die Expression des Zielgens wurde und über Transkriptions- und Translationsanalysen evaluiert. Abschließend wurde die Funktionalität des RMCE anhand eines Kassettenaustauschs in einer optimierten *P. pastoris* RMCE *Master*-Zelllinie nachgewiesen.

Im zweiten Teil dieser Arbeit wird die Entwicklung einer Methode zur sekretorischen Herstellung der Ektodomäne der schwer zu exprimierenden Maus-Transmembranserinprotease Tmprss2 für strukturelle und funktionelle Analysen beschrieben. Tmprss2 spielt eine Schlüsselrolle in der Influenza A Infektion durch die proteolytische Aktivierung von Hämagglutinin H1. Die erste, erfolgreiche Produktion der aktiven Form von Tmprss2 und einer Mutante (D343N) konnte in *P. pastoris* erreicht werden. Die Ausbeuten konnten durch Optimierungen der Kultivierung und des Reinigungsprozesses deutlich erhöht werden. Schließlich wurde die Aktivität des rekombinanten Tmprss2 durch die enzymatische Prozessierung von Hämagglutinin H1 *in vitro* nachgewiesen.

Synopsis

Structural and biochemical analyses usually require high amounts of highly pure target proteins. Therein, the production of the increasingly demanded complex, difficult to express target proteins often presents a major bottleneck. Hence, the optimization of expression hosts and process strategies is an indispensably important task to pave the way for further studies. The methylotrophic yeast *Komagataella phaffii* (commonly known as *Pichia pastoris*) is a widely used expression system for the production of recombinant proteins. It presents an intriguing alternative to the costly mammalian or insect cell culture, as it unites the cost-efficient and simple handling of microorganisms with many of the sophisticated posttranslational processing capabilities of higher eukaryotes. However, the commonly required stable transformation of *P. pastoris* via homologous recombination results in several drawbacks. These include a limited number of established genomic loci for the integration of genes of interest (GOI). Moreoever, due to the strongly varying expression results of the transformants, extensive expression screens to identify high-producer ("jackpot") clones are usually required.

The first part of this work revolves around the establishment of a recombinase mediated cassette exchange (RMCE) system to optimize the expression system *P. pastoris*. The specific integration of GOI into a definedly exchangeable genomic cassette of a *P. pastoris* RMCE master cell line through the enzyme Flp recombinase presents a promising approach to eliminate the expression screening due to comparable productivities of exchanged producer clones. Furthermore, yet uncharacterized genomic loci can be made available for stable protein expression by this system. In this work, the generation of a *P. pastoris* RMCE master cell line is shown. An observed, critical influence of a recognition site of Flp recombinase (FRT F_3) on the expression of the GOI was evaluated through transcriptional and translational analyses. Finally, a proof of concept is presented through a positive RMCE reaction in an optimized *P. pastoris* RMCE master cell line.

The second part of this work deals with the establishment of a method for the secretory production of the ectodomain of the difficult to express transmembrane serine protease mouse Tmprss2 for structural and functional analyses. Tmprss2 plays a key role in influenza A virus infection through proteolytic activation of hemagglutinin H1. The first successful production of both the active form and a mutant form (D343N) of Tmprss2 was achieved in *P. pastoris*. The expression yields could be significantly improved through several optimizations of the cultivation and purification process. Lastly, the activity of recombinant Tmprss2 was proven through its ability to enzymatically process hemagglutinin H1 *in vitro*.

1 Introduction

1.1 The relevance of recombinant protein expression

Proteins are a highly diverse class of complex biological molecules. They are usually built from one or more polymerized chains consisting of the 23 proteinogenic amino acids. Comprising more than 50% of the dry weight of cells, proteins make up the main component of living cells. They participate in many essential functions of living cells taking part in signaling and regulation, for example as transcription factors, enzymes or membrane transporters. (Alberts et al. 2002)

The research of the function of proteins and their regulatory mechanisms is indispensable in order to understand biological processes as a whole, including the research of diseases and infections. For instance, a number of diseases are caused by aberrant behavior of proteins, such as type II diabetes (Pillay and Govender 2013) or Alzheimer's disease (Reitz and Mayeux 2014). Moreover, even normally functioning proteins of a host organism can be exploited by invading pathogens to mediate their infectivity, for example C-X-C chemokine receptor type 4 (CXCR4), which was identified as an essential host co-factor mediating the entry of human immunodeficiency virus into CD4+ T cells (Feng et al. 1996, Arnolds and Spencer 2014). The determination of three-dimensional protein structures at atomic resolution plays a key role in gaining a detailed understanding of protein function. Particularly, the identification of binding partners and functional sites of proteins involved in pathogenesis is a substantial task to facilitate structure based drug design and to develop novel treatment strategies (Buchanan 2002, Zhang and Kim 2003).

One major bottleneck for the determination of protein structures is the generation of sufficient amounts of highly pure, homogenous and soluble protein. However, meeting these requirements is often virtually impossible by protein extraction from natural sources, especially for challenging and complex proteins. For that matter, recombinant protein production is essential to generate the needed target proteins in the desired quality. Another important advantage of recombinant protein expression is the possibility to modify the desired target protein in ways that do not exist in nature.

- The fusion of small peptide tags or whole proteins to the target protein is a widely used approach, for example to simplify its purification by affinity chromatography or to enhance its solubility (Arnau et al. 2006, Esposito and Chatterjee 2006).
- Whole or partial fluorescent marker proteins can also be fused to the target protein to enable direct visual and quantifiable detection of the expression *in vitro* or *in vivo* (Walter et al. 2004, Cabantous et al. 2005).

- Completely or partially customized proteins can be designed, such as alternative antibody fragments such as the single chain variable fragments (scFv) (Carter 2006, Kontermann 2010). Furthermore, it is possible to insert specific mutations into proteins to fit to the desired application, for example to knock out the active site of an enzyme or to mimic the status of a phosphorylation site (Craik et al. 1987, Dissmeyer and Schnittger 2011).
- Finally, in case of difficult-to-express target proteins, the production of partial, truncated protein constructs can significantly enhance the yield of the target protein, which can be a crucial factor for further experiments such as crystallization (Bleckmann et al. 2016a).

The advent of recombinant DNA technology in the 1970s also marked the start for the heterologous expression of recombinant proteins (Cohen et al. 1973), leading to the recombinant production of human insulin in *Escherichia coli (E. coli)* as the first biotechnologically produced pharmaceutical (Goeddel et al. 1979). From there on, an increasing number of recombinant proteins have been produced for several biotechnological and pharmaceutical uses, such as enzymes for biocatalytical applications or vaccines for medical treatment (Makrides 1996, Kumar and Singh 2013).

1.2 Prokaryotic and eukaryotic expression hosts

Since the beginnings of recombinant protein expression, the need for high amounts of new, complex recombinant proteins in high purity increases constantly. Hence, the array of biotechnological expression hosts has since then been extended beyond the popular prokaryotic host *E. coli*. These alternative expression hosts feature several inherent advantages and drawbacks and show significant differences in regard of the expression strategy as well as in protein quality and yield (Hunt 2005).

The gram-negative bacterium *E. coli* is the most abundantly exploited expression host up to now, which is used for the production of high amounts of proteins for pharmaceutical treatments or for structural biology (Nettleship et al. 2010, Kyriakopoulos and Kontoravdi 2013). *E. coli* offers several advantages. As a prokaryotic microorganism, *E. coli* is simple in handling and offers cultivation in inexpensive defined media. It features high growth rates and grows in high cell densities. Moreover, *E. coli* is genetically well characterized. Numerous straightforward and highly optimized protocols have become available over the years. These traits make *E. coli* a very cost- and work-efficient high-throughput system for recombinant protein expression. However, concomitantly with its simplicity, *E. coli* does not possess an endoplasmic reticulum (ER), which is necessary to perform distinct eukaryotic post-translational modifications (Berger et al. 2012). Moreover, *E. coli* is not able to efficiently secrete target proteins into the culture supernatant (Berger et al. 2012). For these reasons, *E. coli* is incapable to facilitate proper folding of many higher eukaryotic proteins (e. g. glycoproteins or membrane proteins), which often results in the

accumulation of misfolded target protein in inclusion bodies (Baneyx and Mujacic 2004, Aricescu et al. 2006, Byrne 2015).

Over 50% of the human proteins are expected to be glycosylated (Apweiler et al. 1999). In these cases the application of eukaryotic cells for heterologous protein production, such as different vertebrate and invertebrate cell lines, can be largely beneficial for protein quality and yield (Wurm 2004). However, cell culture is a significantly more cost- and work-intensive approach in comparison to *E. coli*. Mammalian cell lines are usually applied as expression hosts if human-like glycosylation patterns are desired to ensure full functionality of human target proteins. They offer posttranslational modifications and an array of chaperones, which can be critical to facilitate correct protein folding in order to generate high amounts of soluble target protein. In comparison to *E. coli*, they are characterized by slow growth kinetics in low cell densities. Moreover, they are highly susceptible to mechanical stress and require very specific culture media (Sandig et al. 2005).

Two of the most popular mammalian cell lines are Chinese Hamster Ovary (CHO) cells and Human Embryonic Kidney (HEK) cells. HEK cells comprise one of the most elaborate expression hosts for human proteins. One variant, the HEK293-6E cell line, is particularly suitable for fast transient protein production in suspension cultures. HEK293-6E cells stably express the Epstein Barr virus nuclear antigen 1 (EBNA1), which binds to a replication origin (oriP) within an expression vector, allowing for it to be propagated by replication alongside the chromosomal DNA during mitosis (Van Craenenbroeck et al. 2000, Durocher et al. 2002). By that, the expression vector is retained in the nucleus of the host cells, which allows for a semi-stable protein production over time and concomitantly, higher expression yields. While the EBNA1 system is also available in CHO cell lines (Daramola et al. 2014), these cells are more frequently used for stable protein expression. The gene of interest (GOI) is stably integrated into the genome. This can be performed either randomly or by enzymatic insertion into specific cassette in genetically engineered master cell lines (Wilke et al. 2011, Baser et al. 2016). On the one hand, the generation of stable CHO producer cell lines is a significantly more time-consuming task in comparison to the plasmid based transient expression in HEK cells. On the other hand, the target protein can henceforth be produced continuously. Thus, the scale-up is independent of the addition of expensive plasmid DNA, which can be a critical cost factor.

Alternatively to mammalian cells, several insect cell lines are frequently used as eukaryotic expression systems. Like mammalian systems, insect cells feature an advanced protein folding and secretion system. Hence, they offer full posttranslational modification capabilities, including glycosylation. However, the N-glycosylation pattern that the insect cells produce differs significantly to human cell lines, which presents a major limitation for the pharmaceutical applicability and may influence the full biological activity of the proteins. Insect cells apply a

truncated N-glycosylation pattern compared to mammalian cell lines, most prominently of the paucimannosidic type (Altmann et al. 1999). Unlike the complex type N-glycosylation present in human cells, these glycosylation patterns are less flexible and more homogeneous due to their simpler and shorter structure. This trait makes insect cells a particularly advantageous expression system for protein crystallization, because the proteins usually do not require extensive deglycosylation in order to form protein crystals. For this reason, it is not surprising that insect cells comprise the mostly used eukaryotic expression system for structural biology (PDB January 2017). The most popular insect cell lines used in structural biology are Spodoptera frugiperda SF21 (IPLB-Sf21-AE), its clonal daughter cell line SF9 and Trichoplusia ni BTI 5B1-4 (High Five™, Hi5) (Vaughn et al. 1977, Granados et al. 2007, Durocher and Butler 2009). Compared to mammalian cells, SF21 and Hi5 offer simpler handling and cultivation parameters. They do not require additional support with CO_2 and feature a higher tolerance to both osmolality and by-product concentration (Ikonomou et al. 2003).

The Baculovirus Expression Vector System (BEVS) is frequently used for the expression of recombinant proteins in SF21 and Hi5 cell lines. The BEVS allows for high expression levels through the infection with a recombinant baculovirus harboring an expression cassette for the GOI (Ikonomou et al. 2003). However, the necessary generation of recombinant baculovirus is usually a laborious task compared to the fast transient transfection of HEK cells (Meyer et al. 2013, Dalton and Barton 2014). Furthermore, the baculoviral infection induces proteolysis and cell death, which is unfavorable for continuous or extended expression processes and reliable expression results from batch to batch (Rhee et al. 1999, Dalton and Barton 2014). In addition to that, it is often necessary to work with both SF21 and Hi5 cells for efficient protein production in BEVS. Among these two cell lines, only SF21 cells are capable of producing high amounts of infective baculovirus particles and are hence used for baculovirus amplification (Wilde et al. 2014). Hi5 cells on the other hand exhibit only a low capacity of releasing infective baculovirus, but comprise the superior expression host to produce high amounts of target protein (Wickham et al. 1992, Davis et al. 1993). This can be a critical factor in terms of time consumption and cost efficiency, as it requires the use of different complex media and the handling of both cell lines. Still, the often obtained high protein yields and the homogenous glycosylation pattern usually outweigh these drawbacks making the BEVS a preferred system for high level protein production.

Yeasts like *Pichia pastoris* present a compromise between bacteria and eukaryotic cell lines as expression hosts, displaying a number of advantageous traits of both worlds. They belong to the simplest available eukaryotic expression hosts. As such, they offer full post-translational modification capabilities, while being more comparable to *E. coli* in terms of handling as well as time consumption and cost effectivity. This PhD thesis mainly focuses on the yeast expression host *P. pastoris*, which will be discussed in detail in the following section (§ 1.3). In the end, each of the

eukaryotic and prokaryotic expression systems features several inherent advantages and drawbacks. The choice of the expression host is therefore always dependent on the properties of the desired target protein and the envisioned application.

1.3 The eukaryotic yeast expression system *Pichia pastoris*

The methylotrophic yeast *Pichia pastoris* (*Komagataella phaffii*) is a widely used expression system for the production of recombinant proteins (Ahmad et al. 2014, Bill 2015, Byrne 2015). Apart from that, *P. pastoris* serves as a model organism in cell biology in several studies, including autophagy, peroxisome biogenesis and Golgi function (Subramani et al. 2000, Farre and Subramani 2004, Soderholm et al. 2004). The discovery of the ability of certain yeast species to metabolize methanol as a sole carbon and energy source ranges back to as early as 1969 (Ogata et al. 1969). Methanol presented an interesting carbon source, as it was inexpensive to synthesize from the naturally occurring gas methane. Consequently, the Phillips Petroleum Company (Barthlesville, Oklahoma, USA) made use of this metabolic trait to produce high protein animal feed from yeast biomass (single-cell protein, SCP). They patented the yeast *Pichia pastoris* and established a process protocol to grow it on methanol at high cell densities of greater than 130 g/L dry cell weight. Their progress was halted by the oil crisis in 1973 leading to dramatically increased methane costs. Hence, the SCP protocol never became competitive against another economically more feasible regeneratable protein source (soy beans) (Wegner 1990, Higgins and Cregg 1998).

During the 1980s, the biotechnology company Salk Institute Biotechnology/Industrial Associates, Inc. (SIBIA) (La Jolla, Canada) started to develop *P. pastoris* as a heterologous gene expression system instead, in contract with Phillips Petroleum. At this time, major advancements for the genetic manipulation of *P. pastoris* were achieved at SIBIA, including the establishment of protocols and DNA vectors (Higgins and Cregg 1998). Furthermore, the strictly regulated alcohol oxidase I (*AOX1*) promoter was isolated, which offers strong expression during methanol feed. The *AOX1* promoter in particular synergized well with the media and protocols that Phillips Petroleum had originally developed for the SCP process, resulting in a high yield expression system for heterologous proteins (Higgins and Cregg 1998). The principal system is still essentially in use up to now, with Research Corporation Technologies (Tucson, Arizona, USA) being the present patent holder and Thermo Fisher Scientific (Waltham, Massachusetts, USA) as the licensed reseller.

1.3.1 *Pichia pastoris* as a eukaryotic expression host

As a single celled microbial eukaryote, *P. pastoris* offers several advantages. As a microorganism, *P. pastoris* shares many properties of bacterial expression hosts. It features a high growth rate and the handling of the cells is likewise uncomplicated compared to *E. coli*. *P. pastoris* requires only simple, inexpensive media and grows at moderate temperatures of ca. 28 to 30 °C, similar to the

popular model yeast *Saccharomyces cerevisiae* (*S. cerevisiae*). Additional support of the culture through CO_2 or humidity is not required. These traits present a significant advantage over the cost-intensive and laborious cell culture. Furthermore, *P. pastoris* is easy to genetically manipulate. The commonly applied methods are straightforward with a large and steadily growing array of protocols being available (Lin-Cereghino et al. 2005, Sunga et al. 2008, Looke et al. 2011, Aw and Polizzi 2016). Although it can be forced into mating and diploidy, *P. pastoris* is usually cultivated in its vegetative haploid state (Cregg 1987). Combined with the fact that its genome is completely sequenced (De Schutter et al. 2009, Küberl et al. 2011, Sturmberger et al. 2016), this allows for relatively simple, specific genetic engineering.

While being similar to bacteria in terms of use and handling, *P. pastoris* displays several key advantages over prokaryotic expression systems like *E. coli*. Being a eukaryote, *P. pastoris* offers subcellular compartmentation including an ER. This enables the yeast to natively carry out many posttranslational modifications of higher eukaryotes, such as glycosylation or disulfide bond formation, which can be essential for protein quality and function (Wurm 2004, Byrne 2015). Another contrast to bacterial hosts is presented by the higher capabilities of *P. pastoris* in terms of protein processing and folding (Higgins and Cregg 1998). This includes an unfolded protein response (UPR) system in the ER, which leads to the expression of several stress chaperones to aid protein folding (Guerfal et al. 2010). The UPR system can also be triggered artificially to support the production of soluble, correctly folded recombinant target proteins (Guerfal et al. 2010). Moreover, *P. pastoris* offers powerful secretion capabilities, which allow directing natively secreted heterologous proteins into the culture medium. Notably, *P. pastoris* secretes only few native proteins by itself. This results in the target protein presenting virtually the only protein in the supernatant, simplifying any subsequent purification processes. Since the native secretion sequences of foreign target proteins may not work properly in yeasts like *P. pastoris*, a typical approach is the replacement of the original secretion signal with one compatible to *P. pastoris*. A popular example is the alpha mating factor pre pro secretion signal (MF-α ss) from the baker's yeast *Saccharomyces cerevisiae*. Beneficially, the MF-α ss is cleaved off during secretion by the endogenous Kex and Ste13 proteases of *P. pastoris*. This results in a virtually native N-terminus of the heterologous target protein, although a short sequence of amino acid residues of the MF-α ss can remain at the N-terminus of the protein even after cleavage (Vedvick et al. 1991).

P. pastoris holds two particularly important advantages as an expression host over many other types of yeast including *S. cerevisiae*. First, *P. pastoris* does not tend to perform hyperglycosylation during posttranslational modification, a condition that can even lead in a loss of activity (Eckart and Bussineau 1996, Bretthauer and Castellino 1999). Secondly, *P. pastoris* strictly prefers respiratory growth. Hence, *P. pastoris* does not generate significant amounts of ethanol as a cytotoxic by-product and can therefore be cultivated in very high cell densities (e. g. an OD_{600} of

400 units/mL) (Cregg 2007). This trait presents a considerable advantage over cell culture, as it enables high volumetric product yields. In this regard, *P. pastoris* for example has been reported to outperform CHO cell lines for the production of Fab antibody fragments, although the CHO cells displayed higher "per cell" productivities (Maccani et al. 2014).

All in all, *P. pastoris* combines many of the advantages of the expensive and laborious cell culture and the simple, yet limited bacterial systems for the production of recombinant proteins. As such, *P. pastoris* comprises a highly interesting eukaryotic expression host.

1.3.2 Conventional cell line development in *Pichia pastoris*

For the generation of *P. pastoris* clones expressing the GOI, the principal transformation and selection methods are comparable to those applied in *E. coli* or *S. cerevisiae*. However, the heterologous protein expression in *P. pastoris* commonly relies on stable transformation by homologous recombination (HR), as *P. pastoris* usually does not maintain episomal DNA vectors. Through this, a number of key differences arise for the development of the recombinant clones.

1.3.2.1 *P. pastoris* strains for recombinant protein production

Several genetically engineered *P. pastoris* strains for transformation and recombinant protein expression have been generated over the years, which are all genetically modified derivatives of the wildtype strain Y-11430 from the Northern Regional Research Laboratories (Peoria, Illinois, USA). Apart from antibiotic selection markers, auxotrophic selection based on amino acid biosynthesis is commonly used in *P. pastoris*. For example, the popular strain GS115 (*his4*) was originally generated through random mutagenesis from Y-11430 by Phillips Patroleum. It displays a knockout of the histidine biosynthesis pathway and is not able to grow on minimal medium without the supplementation of histidine (Cregg et al. 1985). Further strains exhibiting different amino acid biosynthesis gene knockouts have been developed likewise. Interestingly, even the commercially offered wildtype strain X-33 (Thermo Fisher Scientific) is principally an engineered strain. X-33 is based on the auxotrophic strain GS115 (*his4*), which was again transformed with the wildtype *His4* gene from *P. pastoris* to restore its prototrophy (Cregg 2007). In addition to auxotrophic mutant strains, several other engineered *P. pastoris* strains have been generated, for example offering deficiencies for endogenous proteases like PEP4. These strains generally allow for enhanced product quality, as these proteases could potentially cleave the target protein (Cereghino and Cregg 2000). Still, the protease knockout strains also display slower growth rates and are more susceptible to stress, for example caused by the transition of the culture to methanol as a sole carbon source (Cereghino and Cregg 2000). This can in turn hamper the expression yield of the target protein (Cereghino and Cregg 2000).

1.3.2.2 Selection of DNA vectors and promoters to drive the target gene expression

In addition to a suitable *P. pastoris* strain, it is important to select an optimal expression vector for protein production. Apart from the selection marker, the choice is primarily dependent on the promoter with the desired properties to drive the expression of the foreign GOI in *P. pastoris*. Common transformation methods in *P. pastoris* are based on stable integration of the expression vector. Consequently, the copy number of the GOI, which can eventually be expected, is several orders of magnitude lower than in transient systems (e.g. in HEK cell lines or in *E. coli*). Thus, the presence of a strong promoter leading to sufficiently high expression results is particularly important. Over the years, several novel promoters have been characterized in *P. pastoris* and numerous attempts have been made to further improve the array of known promoters through genetic engineering (Hartner et al. 2008, Qin et al. 2011, Liang et al. 2013, Periyasamy et al. 2013b, Prielhofer et al. 2013, Zhang et al. 2013). Even the applicability of heterologous promoters, for example from the bacteriophage T7, has been addressed by researchers (Liang et al. 2012). The most prominently used promoters to date in *P. pastoris* for recombinant protein expression are the alcohol oxidase I (*AOX1*) promoter and the glyceraldehyde-3-phosphate dehydrogenase (*GAP*) promoter. Both promoters are part of many commercially available expression vectors, such as the pPICZ and pGAPZ vector families (Thermo Fisher Scientific).

The methanol induced *AOX1* promoter is frequently applied for protein expression in *P. pastoris* by inserting an "*AOX1* promoter – GOI" expression cassette into the genomic *AOX1* locus. *P. pastoris* harbors two native copies of the *AOX* gene designated (*AOX1* and *AOX2*). The *AOX1* promoter is the strongest available native promoter in regard of transcription activity. It is driving 85% of the AOX production, with the *AOX2* promoter being only responsible for the remaining 15% (Cregg et al. 1989). Consequently, the knockout of the *AOX1* gene results in a phenotype that is characterized by slow growth on methanol (methanol utilization slow, mut^S). For example the strain KM71 (*his4 arg4 AOX1::ScARG4*) or its prototroph variant KM71H (*arg4 AOX1::ScARG4*) (Lin-Cereghino and Lin-Cereghino 2007) display this phenotype. In contrast, the unmodified *AOX1* wild type which is present for example in the strain X-33, is referred to as mut^+ (methanol utilization plus). Double knockout strains for *AOX1* and *AOX2* are unable to grow on methanol (mut^-). A particular benefit of using mut^S strains is the significantly lower amount of methanol required to initiate and to maintain the induced expression, which is especially advantageous for scale-up (Cregg 2007).

One key advantage of the *AOX1* promoter for heterologous protein expression is its tight regulation (Cregg 1988, Hartner and Glieder 2006). At non-limiting growth conditions, the transcription of both *AOX1* and *AOX2* is virtually completely repressed in the presence glucose, glycerol and ethanol (Ellis et al. 1985, Cregg et al. 1989, Koutz et al. 1989). By that, the production process can usually be divided into two phases: An initial growth phase using glycerol as a carbon source to

accumulate high cell densities and a discrete production phase by completely switching the carbon source to methanol. This trait can be critically valuable for the production of toxic proteins. Furthermore, the two phase production synergizes well with mut^S strains, often leading to higher productivities (Ascacio-Martinez and Barrera-Saldana 2004, Pla et al. 2006). The fact that mut^S strains grow at slower rates than mut^+ strains under methanol is beneficial in this case, because the cells are directing more resources into protein production instead of cell growth and division.

The major disadvantage of using the *AOX1* promoter is the required use of the highly inflammable and cytotoxic methanol as a substrate, which can be a critical factor for handling and scale-up. The amount of methanol added to the culture for the induction must be tightly regulated, since methanol is converted to formaldehyde as an intermediate product of the methanol utilization pathway. Unlike methanol however, formaldehyde is a cytotoxic substance for *P. pastoris*. If too much methanol is added to the culture, high levels of formaldehyde are produced and subsequently accumulated, because the formaldehyde dehydrogenase is not able to keep up with its further conversion. This eventually causes cell death and is usually detrimental for protein production. To prevent this problem, methanol is only added and maintained at low concentrations of 0.1 to 0.5 % (v/v) to the culture. This particularly complicates small-scale high-throughput expression screenings in comparison to the use of constitutive promoters.

Accordingly, several constitutive promoters have been characterized in *P. pastoris* as alternatives to the methanol dependent *AOX1* promoter (Liang et al. 2013, Periyasamy et al. 2013a, Prielhofer et al. 2013). The most prominent constitutive promoter in *P. pastoris* is the *GAP* promoter, which is commonly applied to drive the expression of recombinant target proteins (Lee et al. 2005, Wen and Wang 2005, Hong et al. 2007). The *GAP* promoter provides a strong transcription rate. The transcriptional activity during feed with glucose is higher than during feed with glycerol or methanol. The main advantage of the *GAP* promoter is that its transcriptional activity is not dependent on the presence of cytotoxic methanol. This is particularly beneficial for large culture volumes and for high-throughput expression screening. In comparison to the *AOX1* promoter, the *GAP* promoter performs about 6-8x weaker for the expression of eGFP (Liang et al. 2013). However, the *GAP* promoter can also outperform the *AOX1* promoter depending on the target protein, as it was reported for the production of alkaline phosphatase (Yang et al. 2015).

In summary, the selection of a promoter and an expression strain greatly depends on the desired application. It influences the cultivation strategy as well as the expected product yield and it can comprise a critical key point for the successful production of the target proteins in high quality.

1.3.2.3 Generation of stable *P. pastoris* producer cell lines by homologous recombination

Method-wise, the generation of *P. pastoris* producer cell is in principal comparable to the protocols usually applied for *E. coli*. First, an expression vector is generated, which is then transformed into *P. pastoris*. Subsequently, positive clones are selected on appropriately selective agar plates. Since the transformation of *P. pastoris* relies on stable integration by HR, at least one homologous region must be present on the expression vector. Commonly, the promoter that drives the expression of the GOI fulfills this role. In contrast to *E. coli* vectors, a *P. pastoris* expression vector has to be linearized inside the homologous region prior to transformation of the cells for the HR to occur at efficient rates. The linearized vector is then transported to the nucleus, where the HR can occur between the homologous region on the vector and the respective genomic locus. The vector is then inserted into the genome and the recombinant cells can be selected for the introduced marker gene. A principal scheme of the generation of recombinant *P. pastoris* clones is depicted in Figure 1.1.

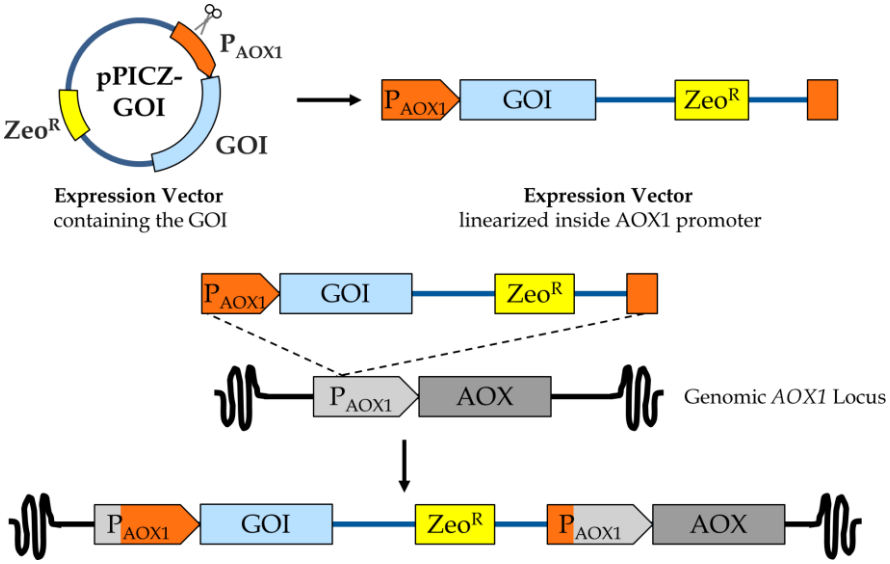


Figure 1.1. Basic principle of the stable transformation of *P. pastoris* by homologous recombination. The expression vector (pPICZ-GOI) harbors the GOI, the *AOX1* promoter (P_{AOX1}) as a homologous region and a Zeocin resistance (Zeo^R). The vector is linearized inside the P_{AOX1} through a uniquely cutting restriction enzyme (displayed by the pair of scissors) and subsequently transformed into *P. pastoris*. Inside the nucleus of the yeast, the homologous recombination (HR) can occur and the vector is stably inserted into the genomic locus homologous to the targeting sequence on vector (the *AOX1* locus). A recombinant producer cell line is generated, which features an expression cassette for the GOI and can be selected for its Zeocin resistance.

In addition to a single vector insertion, the HR can also occur multiple times. Thereby, tandem copies of the vector are inserted into the genomic locus, resulting in multi copy clones. The additional copies of the expression vector can be beneficial for target protein yields. However, this relation is not straightly linear and the expression will decrease again in the presence of too many vector copies (Liu et al. 2014). Even more so, it has been reported that higher expression yields could be achieved in clones displaying lower gene copy numbers for the production of some target proteins, e. g. alkaline phytase (Yang et al. 2012). If two different homologous regions are present at the 5' and 3' ends of the linearized vector, a gene replacement (Ω insertion) can occur at the

genomic region, which is also commonly used not only to introduce a GOI, but also to generate *P. pastoris* knockout strains (Higgins and Cregg 1998). The stable transformation via HR grants several advantages. Once isolated and characterized the recombinant producer clones can be stored and maintained for a virtually indefinite number of production processes with predictable expression results. Additional transformations are not required for later production runs. Furthermore, the expression cultures do not require the permanent addition of selective agents (e. g. antibiotics). This greatly simplifies subsequent handling and scale-up.

However, the use of HR to generate recombinant *P. pastoris* clones also holds a number of drawbacks. While the workflow is initially straightforward in terms of transformation and selection (Figure 1.2), the transformed clones show significant differences in the expression of the gene of interest (GOI). Moreover, the HR machinery of *P. pastoris* is less efficient compared to *S. cerevisiae*, which leads to the risk of off-target integration events through non-homologous end-joining (NHEJ) with unpredictable expression results (Näätsaari et al. 2012). This makes the generation of specifically modified *P. pastoris* strains considerably more complicated (Näätsaari et al. 2012). Interestingly, the expression yield of a target protein is also not solely affected by the inserted copy number of the recombinant expression cassette. Instead, the random insertion of a vector into a favorable genomic locus for the expression of the GOI can also result in significantly higher product yields. Due to the consequently strongly varying expression patterns, extensive screening is usually required to identify a high-producer ("jackpot") clone. This is particularly critical for difficult to express target proteins (Higgins and Cregg 1998). Hence, the generation of a producer clone with suitable expression yields can result in a time-consuming and work-intensive process (Brooks et al. 2013).

With every transformation, the expression vector is stably integrated into the genome, thereby permanently reserving its inherent selection marker gene from further use. This heavily complicates the process of introducing multiple genetic modifications to the yeast, because the range of antibiotic selection markers in *P. pastoris* is limited (Wan et al. 2004, Papakonstantinou et al. 2009). Hence, recycling of an antibiotic selection gene after a transformation is of great interest especially for genomic engineering applications. However, marker gene recycling usually presents a complicated and laborious procedure in itself, requiring site specific recombinases (Flp) or HR to loop out the selective cassette and remove it from the genome (Cregg and Madden 1989).

The use of additional genetic elements such as negative selection markers is largely beneficial to efficiently select the resulting clones, yet this further complicates the process (Yang et al. 2009). Several attempts have been made to solve these various shortcomings of the stable cell line generation by HR in *P. pastoris*, for example by increasing HR efficiencies through the deletion of the gene encoding the KU70 homolog of *P. pastoris*. KU70 is a known key-player in the non-

homologous-endjoining repair system (NHEJ) in *S. cerevisiae*. Its deletion indeed led to a significant increase of the HR targeting efficiencies in *P. pastoris*, but it also resulted in vastly decreased transformation efficiencies, reduced growth rates and higher susceptibility to random mutations of the genome (Näätsaari et al. 2012).

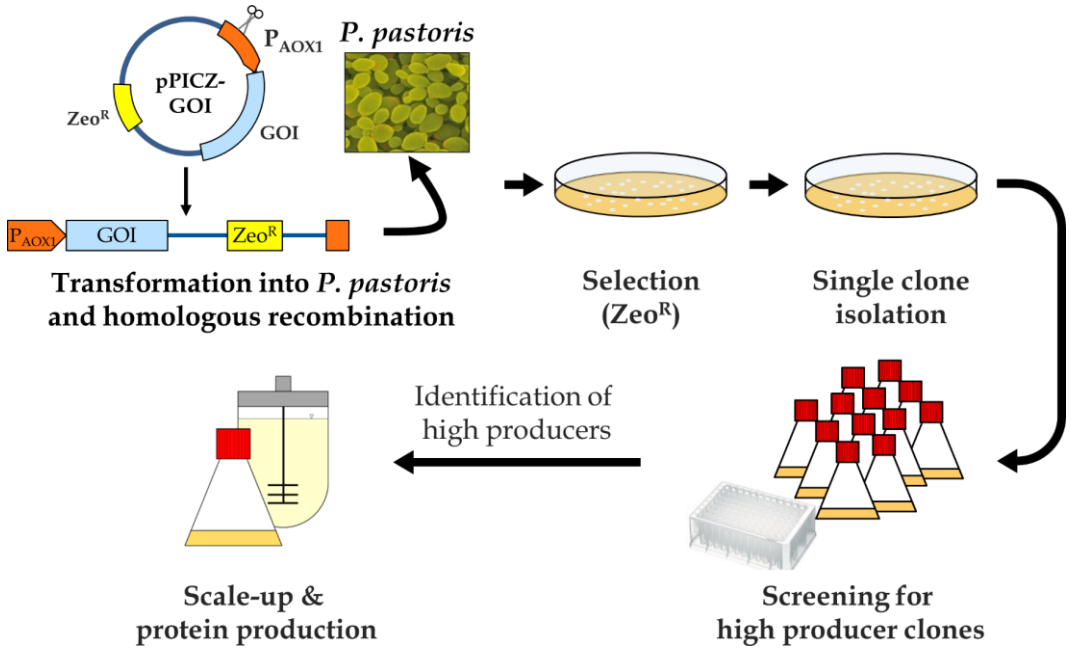


Figure 1.2. Schematic overview of the principal work flow for protein expression in *P. pastoris*.
An expression vector harboring the GOI (pPICZ-GOI) is linearized and transformed into *P. pastoris*. Following primary selection (e.g. for Zeocin resistance, Zeo^R) (~3 days), grown colonies are singly streaked out for clone purification /single clone isolation (~3 days). To isolate multi copy clones, the single clone isolation step includes screening for growth on an increasing concentration gradient of the selective agent (e. g. Zeocin). The isolated clones are tested for their expression levels, which can take up several days or weeks, depending on the target protein. The identified high producer clones can then be used for scaled-up production processes.

1.3.2.4 Episomal vector propagation by PARS sequences

Vectors containing a Pichia Autonomous Replicating Sequences (PARS) comprise an alternative to the commonly applied stable integration of expression vectors, as they allow for a limited, episomal propagation of plasmids in *P. pastoris* (Cregg et al. 1985). Originally, ARS sequences were first discovered for *S. cerevisiae* (Szabo et al. 2003), where they are believed to play a role as DNA origins of replication (ori) during mitosis (Huberman et al. 1988).

In contrast to the stable transformation via HR in *P. pastoris*, a vector containing a PARS sequence can be efficiently transformed as a circular plasmid without the need for linearization, thereby simplifying the procedure (Cregg et al. 1985, Lee et al. 2005). Furthermore, PARS vectors like pBGP1 are only kept at very low copy numbers inside the cells, averaging around one copy per cell. For these reasons, they have been applied for fast transient expression screenings as a time-saving alternative to stable cell line generation (Aoki et al. 2003, Lee et al. 2005). However, the low copy number usually results in consequently low expression levels. Moreover, the presence of homologous sequences such as promoters or terminators on PARS plasmids can cause their

uncontrolled integration into the genome of *P. pastoris* after a random period of time (Higgins and Cregg 1998). By that, an originally isolated, homogeneous clone may diverge into a heterogeneous mixture of different clones, which can significantly compromise the expression levels. Therefore, it is not surprising that the majority of expression screens rely on stable transformation (Macauley-Patrick et al. 2005), as only stably transformed clones offer long-term stability and constant product yields from an identified "jackpot" clone.

1.4 The multi-host expression system

The multi host expression system (mHost-XS) is a versatile system for protein expression screening and production in an array of different expression hosts. The basis of the mHost-XS is the unique single donor vector pFlpBtM (Meyer et al. 2013). It offers all the elements needed for fast transient expression in HEK293-6E cells, baculoviral-driven expression in insect cells as well as stable expression in CHO master cell lines via recombinase-mediated cassette exchange (RMCE). A basic scheme of the vector is depicted in Figure 1.3.

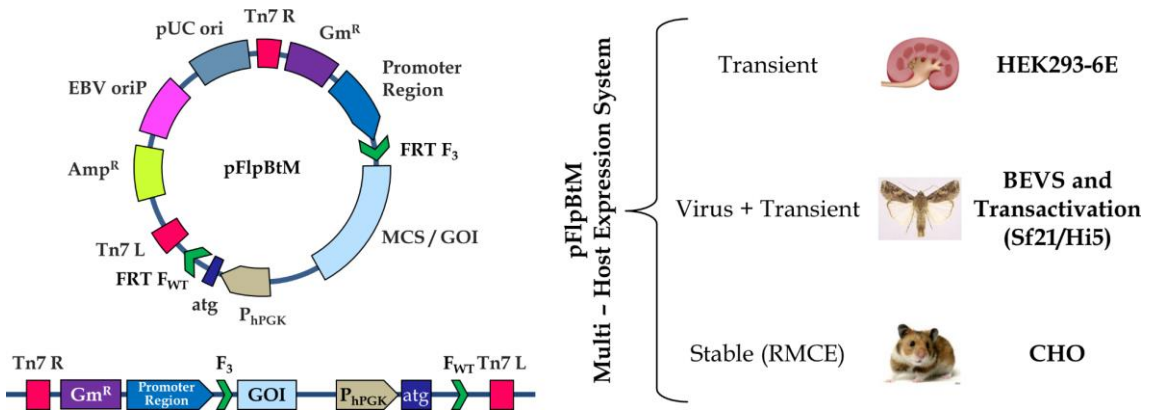


Figure 1.3. Overview on the versatile multi-host vector pFlpBtM.
The vector pFlpBtM contains all necessary elements for fast plasmid-based expression in HEK293-6E cells. The promoter region includes the strong CMV promoter and the EBV oriP mediates the replication of the plasmid in the host cells during mitosis. Furthermore, pFlpBtM can be used to generate recombinant bacmids for the BEVS via Tn7 transposition. Lastly, pFlpBtM can serve as a donor vector for a Flp recombinase based RMCE, due to the presence of a pair of heterospecific FRT sites (FRT F_3 and F_{WT}). This system is applied to generate stable CHO producer cell lines. A human PGK promoter (hPGK) and an ATG start codon (atg) are present inside the RMCE cassette to complement a selection trap in these cell lines.

The expression of both intracellular and secretory proteins is possible with pFlpBtM. The secretion is realized by an N-terminal IgG secretion signal. Furthermore, pFlpBtM features C-terminal tags for protein purification and a TEV protease cleavage site to cleave them off. For the transient expression in HEK293-6E cells, the pFlpBtM comprises the strong cytomegalovirus (CMV) promoter. Furthermore, HEK293-6E cells constitutively express the Epstein-Barr nuclear antigen 1 (EBNA1). EBNA1 mediates the replication of pFlpBtM and its propagation to the daughter cells during mitosis through an Epstein-Barr virus (EBV) oriP, which is present on the vector backbone.

The pFlpBtM plasmid further harbors all necessary elements for the Bac-to-Bac® baculovirus expression vector system (BEVS) (Thermo Fisher Scientific). The promoter region of pFlpBtM harbors baculoviral promoters for the expression of the GOI. For the generation of the recombinant bacmid, a specific *E. coli* strain is necessary, designated as DH10Bac. It carries a recombinant bacmid (EmBacY) and a helper plasmid coding for the enzyme transposase (Trowitzsch et al. 2010). When DH10Bac is transformed with pFlpBtM, the transposase mediates the defined integration of the vector region between the two Tn7 sites (Tn7 L and Tn7 R) present on pFlpBtM into the mini-attTn7 site on the EmBacY. The mini-attTn7 site is embedded into the coding sequence of a lacZ gene in EmBacY. Through the transposition, the lacZ gene is disrupted, enabling visual selection for positive integration events by blue-white screening. Once the bacmid is purified, it is transfected into Sf21 insect cells where it causes the production of recombinant baculovirus particles. The transfection with the bacmid as well as any subsequent infection with the recombinant virus can be monitored through the expression of YFP, which is present on the bacmid backbone. A schematic overview of the workflow is shown in Figure 1.4.

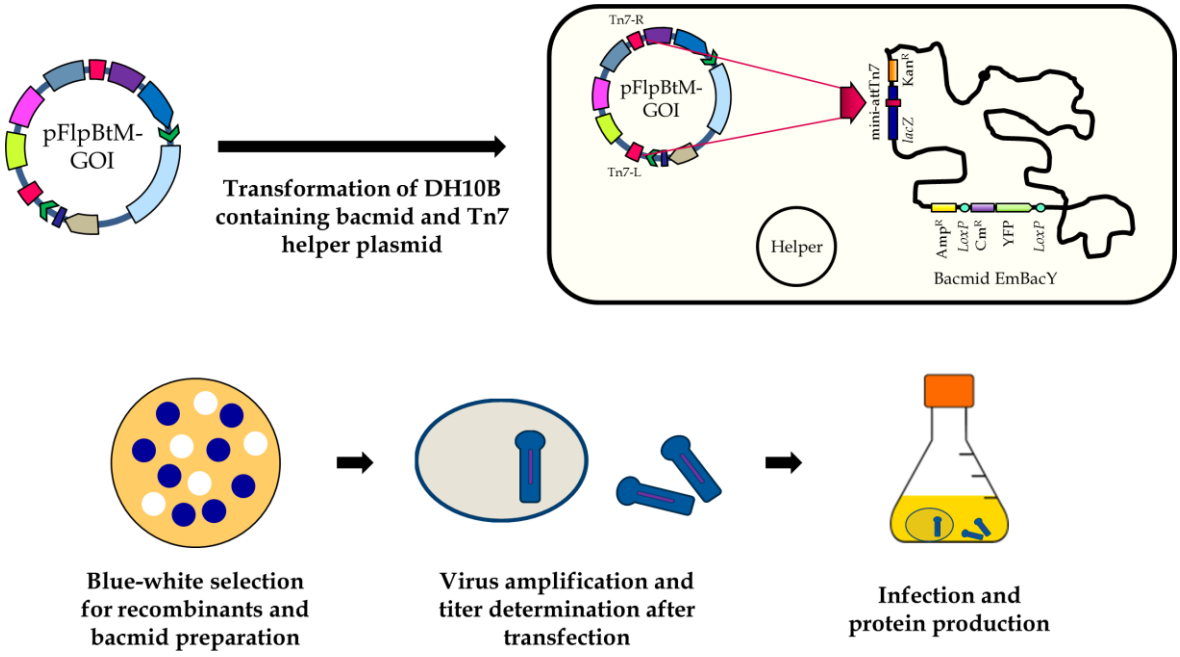


Figure 1.4. Schematic workflow of the generation of baculovirus for BEVS using the mHost-XS.
The vector pFlpBtM, which harbors the GOI, is transformed into *E. coli* DH10Bac carrying the EmBacY bacmid. The transposase coded by a helper plasmid mediates the transposition of the vector region of pFlpBtM between the Tn7 sites into the corresponding mini-attTn7 site on the bacmid. The mini-attTn7 site is embedded within a *lacZ* gene, which is consequently disrupted, enabling blue-white screening. After purification, the recombinant bacmid can be used for the generation of baculoviruses and for the infection of insect cells. A YFP that is coded from the bacmid backbone of EmBacY aids monitoring the viral infection.

In addition to this, insect cells can be brought to recombinant protein expression in a process called transactivation. This is performed by transiently transfecting of the insect cells with pFlpBtM (containing the GOI) and infecting them with "empty" baculovirus, which produces the necessary

proteins and transcription factors to induce protein expression directly from the plasmid (Bleckmann et al. 2016b).

Finally, pFlpBtM serves as a donor vector for the stable transformation of specific CHO RMCE master cell lines via a recombinase-mediated cassette exchange (RMCE). These cell lines are stably transformed ("tagged") with an exchangeable RMCE cassette. The expression of the exchangeable GOI is driven by an upstream residing "endogenous" EF promoter that is supplied by the transformed tagging vector. Upon co-transfection with pFlpBtM and a helper vector for the expression of Flp recombinase, the genomic RMCE cassette is exchanged for its counterpart on pFlpBtM. Thereby, the GOI and additional elements for selection of positive integration events are inserted into the genome. An overview of the RMCE method is presented in the following section.

1.5 The recombinase mediated cassette exchange (RMCE)

The RMCE system was originally established by Schlake and Bode in 1994 (Schlake and Bode 1994). It is a convenient method for the highly specific and robust stable insertion of DNA cassettes into a predefined genomic locus of a host cell using the Flp recombinase (flippase, Flp), a member of the tyrosine family of site-specific recombinases (Ma et al. 2007). The RMCE is based on the generation of stable master cell lines (MCL) in a process called tagging, in which an exchangeable RMCE cassette is stably integrated into the genome of the desired expression host. It comprises a number of different elements, for example a fluorescent marker and/or an antibiotic resistance gene for robust primary selection. The promoter to drive the expression of the fluorescent marker (or a GOI after a cassette exchange) can either reside stably upstream of the tagging cassette inside the genome or it can be part of the exchangeable RMCE cassette. The borders of the exchangeable RMCE cassette are defined by a pair of heterospecific FRT (Flp recombinase targeting) sites in parallel orientation, for example FRT F_3 and FRT F_{WT}. FRT sites comprise short repeating 13 bp DNA sequences comprising two direct repeats and an inverted repeat separated by a variable 8 bp spacer (Schlake and Bode 1994). They form secondary structures, which are recognized by the Flp recombinase. The RMCE is based on the fact that the Flp is highly specific and only allows for recombination between identical FRT sites. Thus, the different FRT site variants that were obtained by mutating the spacer region do not display cross-reactivity and are only recombined with an identical counterpart (Schlake and Bode 1994).

The principle of the RMCE reaction is shown in Figure 1.5. To perform the RMCE, an identified stable RMCE MCL, a suitable donor vector and a helper vector delivering the Flp recombinase are needed. The donor vector (e. g. pFlpBtM) has to harbor an exchange cassette flanked by the same pair of FRT sites that is also flanking the genomic tagging cassette. When the donor vector and the

helper vector are transformed into the RMCE MCL, the RMCE reaction can occur following the expression of Flp recombinase.

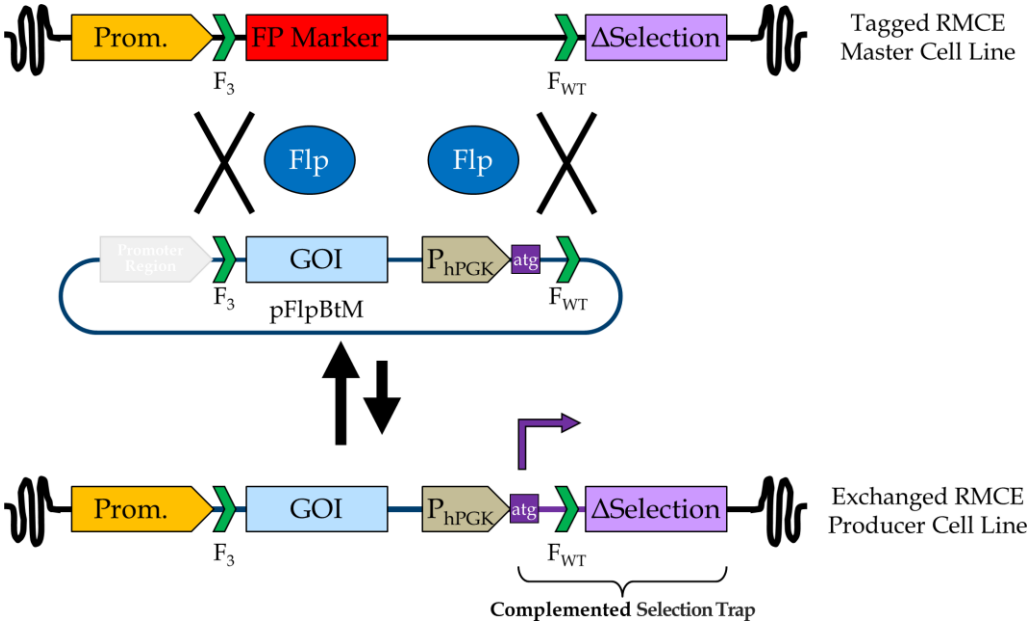


Figure 1.5. Schematic depiction of the RMCE reaction using the mHost-XS vector pFlpBtM.
In the presence of Flp recombinase, the RMCE cassette of the donor vector (pFlpBtM) flanked by FRT F_3 and F_{WT} is exchanged with a parallel genomic tagging cassette, featuring an identical pair of FRT sites. Thereby, the GOI is stably inserted into the genome. The promoter to drive its expression is for example present "endogenously" in the genomic tagging locus ("Prom."). The promoter region of pFlpBtM is irrelevant to the reaction, as it is not inserted into the genome (greyed out). Next to the GOI, a human PGK promoter (hPGK) and an ATG start codon are delivered to complement an otherwise inactive selection gene (selection trap, ΔSelection). The exchanged clones are screened for the selectivity conveyed by the complemented selection trap. This vastly improves screening for positively exchanged producer clones, as the excision of the donor cassette is generally thermodynamically favored over its insertion.

Flp recombinase uses a type IB topoisomerase like mechanism to carry out the cassette exchange in a repeated two-step reaction process (Ma et al. 2007). It comprises first the creation of a Holliday junction intermediate and secondly the recombination of two complementary DNA strands presented by a pair of identical FRT sites (Ma et al. 2007). Therein, the cleavage of the DNA strands is performed by a highly conserved tyrosine nucleophile (Ma et al. 2007). Through the recombination of both FRT sites flanking the RMCE tagging cassette with their respective counterparts on the donor vector, the exchangeable donor cassette is inserted into the genome. Consequently, the elements of the tagging cassette are removed from the genomic locus. The Flp mediated cassette exchange is a dynamically reversible reaction process, with the excision reaction being thermodynamically favored over the insertion reaction (O'Gorman et al. 1991, Schübeler et al. 1998, Francastel et al. 1999). Hence, the application of selection markers for a successful integration of the donor vector into the genome is greatly beneficial to identify successfully exchanged clones. One particularly useful approach is to place a so-called selection trap directly downstream of the RMCE cassette. The selection trap presents an inactive selection marker gene, which lacks its ATG start codon and a promoter to induce its expression. Both of these elements

are delivered by the donor vector during a RMCE, thereby complementing the selection trap gene. The cells can subsequently be screened for the newly acquired selectivity. In contrast to the delivery of a full selection marker cassette, the selection trap method only provides the selectivity upon the defined vector integration via RMCE, avoiding false positive clones from random vector insertions into the genome. By that, the selection trap provides an efficient method to robustly select for positive RMCE events (Verhoeyen et al. 2001, Wilke et al. 2011, Baser et al. 2016).

Anticipated benefits of establishing the RMCE system in *P. pastoris*

The application of the RMCE system in *P. pastoris* displays several anticipated advantages over the current system based on HR. First, the RMCE should result in a highly specific and robust exchange of a GOI into a defined genomic locus, circumventing the problems arising from the use of HR.

Secondly, the tagging with the RMCE cassette could be intentionally performed randomly to identify yet uncharacterized genomic loci for stable recombinant protein expression. Thereby, the established loci (*AOX1* and *GAP*) could be left untouched for further potential applications.

Thirdly, antibiotic selection markers can be placed inside the tagging RMCE cassette for robust primary selection during the generation of the MCL. The RMCE reaction would directly recycle these selection markers without any additional procedures.

Lastly, the expression for the production of the target protein would only have to be screened once through the expression of a fluorescent marker protein from the RMCE tagging cassette of the MCL clones (as a model GOI). After the RMCE, the expression level of the integrated GOI should scale with that of the exchanged fluorescent marker. In other words, the RMCE is expected to result in clones with uniform expression results. This would eliminate the usually required extensive expression screens.

1.6 Fluorescent proteins as visual expression markers

Fluorescent proteins such as the green fluorescent protein (GFP) from *Aequorea victoria* (Shimomura et al. 1962) are commonly applied as marker proteins in many studies including gene expression and protein-protein interaction, as they allow for highly sensitive, straightforward and quantifiable measurements (Chalfie et al. 1994, Walter et al. 2004, Cabantous et al. 2005). GFP and other GFP-like fluorophores are usually simple proteins. This enables their heterologous production even in bacterial expression hosts like *E. coli* (Prasher et al. 1992). Moreover, the chromophore is formed intrinsically through an extended π-electron system by conserved amino acid residues of the protein (e.g. Ser^{65}-Tyr^{66}-Gly^{67} for GFP) (Ormö et al. 1996). Hence, the measurements of the

respective fluorescence do not require the addition costly substrates (Chalfie et al. 1994, Chalfie 1995, Rodrigues et al. 2001).

For these reasons, fluorescent proteins were selected as marker proteins to establish the RMCE system in *P. pastoris* in this PhD study. It should be noted that yeasts like *P. pastoris* intrinsically display a weak auto fluorescence, which particularly interferes with GFP measurements (Cormack et al. 1997, Hitchcock et al. 2006). In contrast, the measurement of red fluorescent proteins (RFP), which is performed at longer wavelengths, greatly reduces the influence of both light scattering and the intrinsic auto fluorescence of yeast cells (Hitchcock et al. 2006). An additional advantage of using RFPs as markers for gene expression is the fact that the expression is already visible at ambient light. Therefore, two RFP were selected as fluorescent markers to track the tagging of the yeast cells with the RMCE cassette: The monomeric RFP mCherry from *Discosoma sp.*, which offers a particularly high photostability (Shaner et al. 2004), and the RFP from *Corynactis californica*, which is available as "RudolphRFP" in a codon optimized form for high level expression in *P. pastoris* (ATUM, Newark, California, USA).

It had been shown in prior studies with CHO RMCE master cell lines that the excitation and emission wavelengths of green and red fluorescent proteins differ sufficiently to enable separate measurements of both chromophores from the same cell (Baser et al. 2016). Hence, GFP was chosen as a model GOI to monitor the cassette exchange reaction in *P. pastoris*. As mentioned before, the auto fluorescence of *P. pastoris* interferes with GFP measurements. For this reason, a suitable variant of GFP was selected to counter said interference and to generate a sufficiently strong signal in *P. pastoris*. This variant is based on enhanced GFP (eGFP), which is a mutant form with significantly higher illumination levels (Cormack et al. 1996). For high level expression in yeast cells specifically, the *eGFP* gene was further codon optimized (yeast enhanced GFP; y-eGFP) (Cormack et al. 1997). A synthetic clone of y-eGFP generated by Christian Kambach (HZI, Braunschweig) is used in this work. The parameters of the fluorochromes used in this PhD study are listed in Table 1.1.

Table 1.1. Properties of the fluorescent marker proteins used in this work.

Name	Excitation [nm]	Emission [nm]	MW [kDa]	Length [aa]	Reference
mCherry	587	610	26.7	236	Shaner et al. 2004
RudolphRFP	553	570	25.3	226	ATUM (Newark, CF, USA)
y-eGFP	484	510	26.9	238	Christian Kambach, HZI based on Cormack et al. 1997

1.7 The type II transmembrane serine protease Tmprss2

The major portion of the proteases in the respiratory tract consists of serine proteases (Meyer and Jaspers 2015). Among these is the family of the type II transmembrane serine proteases (TTSP), which display a number of common structural traits. They comprise an N-terminal intercellular tail, followed by the transmembrane domain and an ectodomain containing a variable stem region and the C-terminal serine protease domain (SPD) (Bugge et al. 2009). TTSPs are initially expressed as single-chain proenzymes, which require auto-catalytic activation by cleavage of a conserved zymogenic site downstream of a basic arginine or lysine (Bugge et al. 2009). The zymogenic activation results in molecular rearrangement of the ectodomain, while the cleaved parts of the protein are covalently kept together by a conserved disulfide bond (Bugge et al. 2009).

TTSPs are divided in four subfamilies based on phylogenetic analyses of their SPD and their domain structure (Figure 1.6). These are the Matriptase subfamily, the HAT / DESC (human airway trypsin-like protease / differentially expressed in squamous cell carcinoma) subfamily, the Hepsin / Tmprss (transmembrane protease serine) and finally the Corin subfamily (Netzel-Arnett et al. 2003, Szabo et al. 2003, Szabo and Bugge 2008, Antalis et al. 2010).

Tmprss2 is a member of the Hepsin/Tmprss subfamily. Similar to other TTSPs such as Matriptase or HAT, Tmprss2 is widely expressed in epithelial tissues, including the prostate epithelium, the ovaries as well as the upper airway epithelium and the alveoli (Lin et al. 1999, Bugge et al. 2009). Depending on its localization and tissue site, Tmprss2 can exist as either a full length protein or in a variety of truncated forms (Afar et al. 2001, Chen et al. 2010). Apart from the membrane-bound state, the ectodomain can also be released into the pericellular environment (Yasuoka et al. 1997, Afar et al. 2001). Tmprss2 was shown to reduce epithelial sodium channel activity, which led to the proposition that Tmprss2 might be responsible for regulating ion transport and sodium currents in the airway epithelium (Donaldson et al. 2002). Moreover, Tmprss2 reportedly takes part in inflammatory responses in the prostate by proteolytic activation of protease-activated receptor 2 (Wilson et al. 2005). Yet, the exact physiological function of Tmprss2 is still unclear, since Tmprss2 knockout mice were healthy and showed neither a discernible phenotype nor any compensatory up-regulation of other TTSPs (Kim et al. 2006).

According to different reports, Tmprss2 is speculated to play a key role in the activation and subsequent replication of respiratory viruses, such as the H1N1 influenza A virus, the human metapneumovirus or the SARS coronavirus (Böttcher et al. 2006, Shirogane et al. 2008, Böttcher et al. 2009, Bertram et al. 2012).

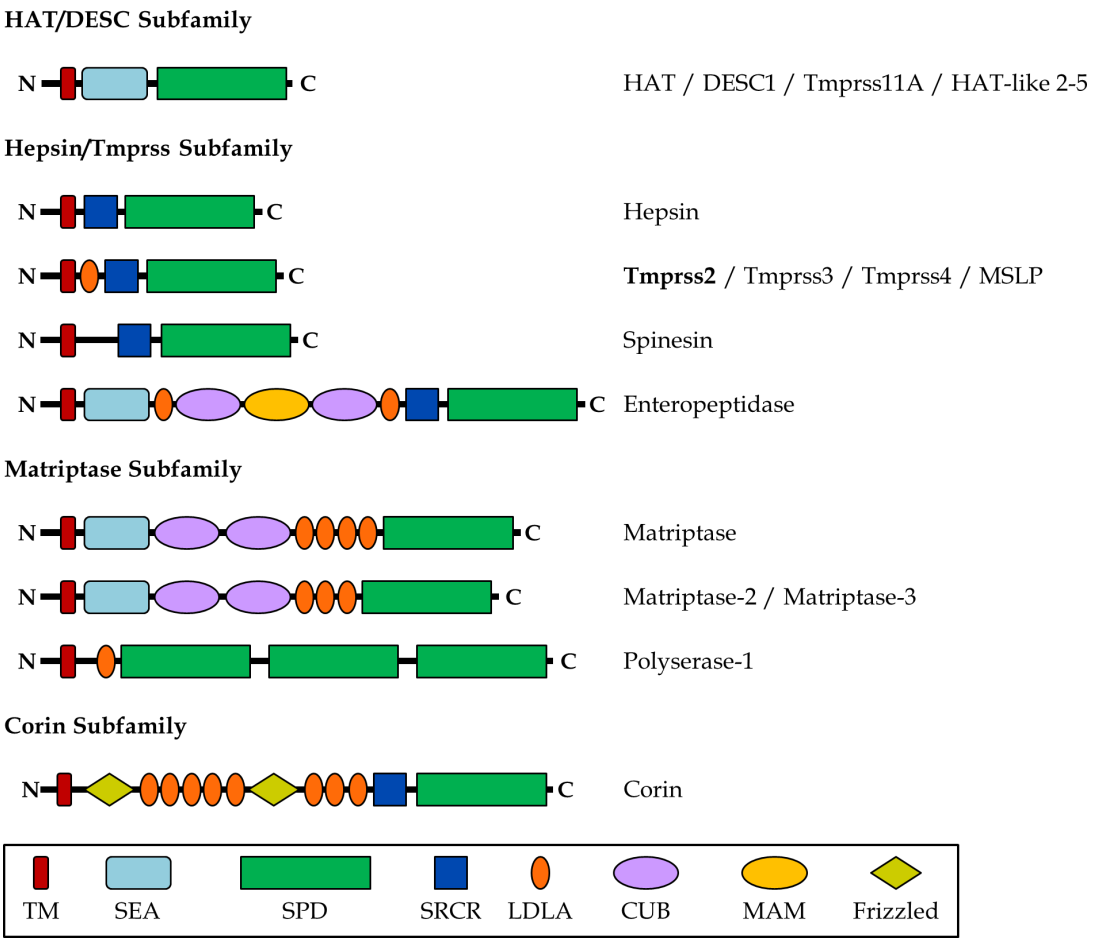


Figure 1.6. Simplified schematic overview of the TTSP subfamilies.
TTSPs are divided into four subfamilies: The HAT/DESC subfamily, the Hepsin/Tmprss subfamily, the Matriptase subfamily and the Corin subfamily. A schematic view of the domain structures is depicted on the left-hand side. The names of the respective members are depicted on the right-hand side. Tmprss2 (bold) is a member of the Hepsin/Tmprss subfamily. For simplification, the emphasis was laid on the basic domain structures. Differences in the exact lengths of the N- and C-terminal tails are not depicted. All TTSPs are zymogens, which require autocatalytic activation through cleavage upstream of the respective serine protease domain. **Abbreviations:** HAT: human airway trypsin-like protease; DESC: differentially expressed in squamous cell carcinoma; Tmprss: transmembrane protease serine; MSLP: mosaic serine protease large-form. TM: transmembrane domain; SEA: sea urchin sperm protein/enteropeptidase/agrin domain; SPD: serine protease domain; SRCR: scavenger receptor cysteine-rich domain; LDLA: low density lipoprotein receptor class A; CUB: Cls/Clr, urchin embryonic growth factor bone morphogenetic protein-1 domain; MAM: meprin/A5 antigen/receptor protein phosphatase mu domain; Frizzled: Frizzled domain.

1.7.1 The role of Tmprss2 during influenza A pathogenesis

Seasonally occurring influenza epidemics as well as unpredictable pandemics are a grave threat to public health. This is exemplified by the 1918 flu pandemic caused by the H1N1 influenza A virus (IAV). It infected over 500 million people world-wide, resulting in approximately 50 million deaths (Taubenberger and Morens 2006). Current therapy approaches rely on targeting the viral proteins neuraminidase and M2. However, the development of resistance to these anti-viral drugs is frequently observed, because of the high mutation rate of the virus (Hurt et al. 2012). For this reason, the identification of new drug targets for the treatment of influenza A infection is an essential task. Therein, invariable host cell factors, which are critically essential for viral

pathogenesis but not for cellular survival, comprise highly interesting targets to develop novel strategies for antiviral therapy (Hatesuer et al. 2013).

The activation of the viral surface protein hemagglutinin (HA) by proteolytic cleavage of its precursor form (HA0) into its disulfide-linked subunits HA1 and HA2 presents an essential step for IAV infectivity (Klenk and Rott 1988, Steinhauer 1999). The majority of the IAV subtypes harbor HAs with monobasic cleavage sites, which are only cleaved in a limited number of tissues such as the respiratory tract (Klenk and Rott 1988, Steinhauer 1999). The TTSPs Tmprss2 and HAT, which are present on airway epithelial cells, were identified as possible candidates to catalyze this reaction and subsequently shown to be able to proteolytically activate the monobasic HAs H1, H2 and H3 in MDCK cell culture, resulting in enhanced viral replication (Böttcher et al. 2006, Böttcher et al. 2009). Consistently with these results, Tmprss2 was later found to be critically essential for the activation of H1N1 and - to a lesser extent - H3N2 IAV in living mice, as Tmprss2$^{-/-}$ knockout mice were significantly less susceptible to H1N1 infection compared to wildtype mice (Hatesuer et al. 2013). Interestingly, Tmprss2 deficient mice were in contrast not protected from infection by a multi-basic H7N7 IAV (Hatesuer et al. 2013). These results were further supported by different reports, which also showed that Tmprss2 deficient mice were protected from H1N1 infection (Sakai et al. 2014, Tarnow et al. 2014).

As mentioned before, Tmprss2 knockout mice do not show any phenotypic alterations (Kim et al. 2006). This makes Tmprss2 an interesting drug target to develop novel treatment strategies against IAV infection by structure-based drug design (SBDD). Solving the crystal structure of mouse Tmprss2, especially of the enzyme-substrate interaction surface in combination with HA H1, presents an important step on the path to gain a deeper understanding the proteolytic activation of HA by Tmprss2. This PhD thesis deals with the establishment of a method to generate the required amounts of mouse Tmprss2 for crystallization experiments.

1.7.2 Mouse Tmprss2 as a recombinant target protein

Mouse Tmprss2 (NCBI reference number NP_056590) is a 490 aa long protein with a molecular weight of ~70 kD. Its extracellular domain comprises a low density lipoprotein receptor class A domain (LDLA), a scavenger receptor cysteine-rich domain (SRCR) and the catalytic serine protease domain (SPD). The total extracellular domain spans over 384 aa with a molecular weight of ~45 kD, with the C-terminal serine protease domain taking up 236 aa (~30 kD). A schematic overview of the protein domain structure of mouse Tmprss2 is depicted in Figure 1.7.

The selection of a suitable construct and an expression host is essential to gain the required amounts of mouse Tmprss2 for structural analyses. The SPD carries out the primary proteolytic activity of Tmprss2 and thus comprises an essential part of the recombinant construct for SBDD.

The LDLA and SRCR domains are speculated to play an important role for the activity of Tmprss2, as the proteolytic activity was significantly reduced for soluble truncated fractions comprising only the C-terminal SPD (Afar et al. 2001, Böttcher-Friebertshauser et al. 2010). Similar observations were made for another TTSP, Matriptase-2. Matriptase-2 harbors a SEA domain and four LDLA domains (LDLA1-4) upstream of its SPD (Figure 1.6). Blocking LDLA3 or mutating either LDLA2 or SEA reportedly hindered zymogenic activation and resulted in a loss of function (Lee et al. 2007, Ramsay et al. 2009, Silvestri et al. 2009). Furthermore, despite the homologies to other TTSPs, only Tmprss2 can cleave HA H1 efficiently *in vivo*, as evident by the previously mentioned experiments carried out in living mice (Hatesuer et al. 2013). Hence, the LDLA and SRCR domains of Tmprss2 might play an important role in the activation and substrate binding capabilities of Tmprss2. Consequently, the full length ectodomain of Tmprss2 was selected to serve as the recombinant target protein for SBDD.

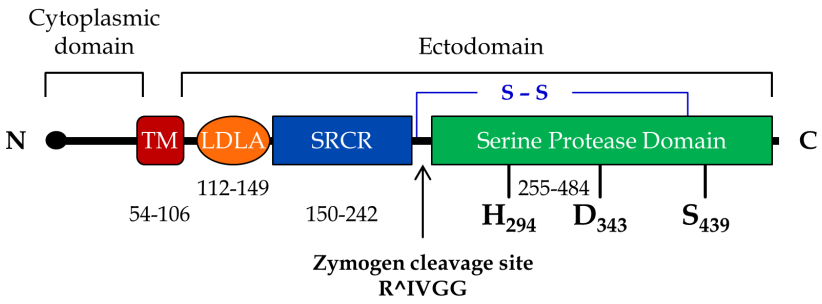


Figure 1.7. Schematic overview of mouse Tmprss2.
Mouse Tmprss2 comprises 490 amino acids. The N-terminal, cytoplasmic domain is followed by a transmembrane domain (TM) and the ectodomain, which comprises a LDLA (low density lipoprotein receptor class A), a SRCR (scavenger receptor cysteine-rich) and a serine protease domain. The conserved residues of the catalytic triad were identified as histidine 294 (H294), aspartate 343 (D343) and serine 439 (S439) (Vaarala et al. 2001). The zymogenic cleavage site is located between the SRCR and the serine protease domains. A covalent disulfide bond (S – S) keeps the protein attached following autocatalytic activation.

Comparable to other TTSPs, Tmprss2 is a multi-domain glycoprotein. Tmprss2 harbors two cysteine-rich domains (LDLA and SRCR) with multiple disulfide bonds. It is conceivable that the folding machinery of eukaryotic cells and their ability to perform posttranslational modifications in the ER is largely beneficial for the production of the ectodomain of Tmprss2. In accordance to this, several groups reportedly relied on *P. pastoris* for the secretory production of the full ectodomains of other TTSPs (Hepsin, DESC1 and Matriptase-1), as well as the SPD of Matriptase-2 (Friedrich et al. 2002, Somoza et al. 2003, Yuan et al. 2011, Zhao et al. 2013). Moreover, the ectodomains of Matriptase-2, Hepsin and DESC1 were also expressed in Drosophila S2 insect cells (Beliveau et al. 2009). It should be noted that the same research group only reported the production of the SPD of Matriptase-1 in *E. coli* (Désilets et al. 2006, Beliveau et al. 2009). Taken together, the reports indicate that eukaryotic cells likely present the most suitable expression system for the whole ectodomain of a TTSP. For this reason, *P. pastoris* was selected as the initial expression host to test the production of the ectodomain of mouse Tmprss2 in this PhD thesis.

1.8 Aim of this work

As described in the introduction, *Pichia pastoris* is a powerful eukaryotic expression host for the cost-efficient production of recombinant proteins. Still, *P. pastoris* displays some drawbacks. The number of established genomic loci for stable protein expression is still limited. And even after targeting an expression vector into an established locus (*AOX1* or *GAP*) by homologous recombination, extensive screening is required to identify high producer ("jackpot") clones. The expression levels are not solely based on gene copy number, but also on random integration events.

Within the context of protein expression, two general aims exist. On the one hand, the time-efficient generation of a jackpot clone is desired for any gene of interest. Hence, the efficiency of this process has to be improved. On the other hand, it is also of interest to optimize an expression system to enhance the production of a specific target protein.

For the first aim, this PhD thesis deals with the improvement of the expression system *P. pastoris* through the establishment of a recombinase mediated cassette exchange (RMCE) system to specifically integrate a GOI into any genomic locus harboring a stable RMCE cassette. This should result in the fast generation of producer clones with uniform product yields. The exchange of a GOI into a jackpot clone should likewise result in a high producer clone for the GOI. In this work, an initial pilot for a *P. pastoris* RMCE system should be established, using the genomic *GAP* locus to ensure comparability of different tested variants of the RMCE cassette. The RMCE cassettes should be designed to meet the specific demands of *P. pastoris*, including an appropriate selection trap to select for exchanged clones. A set of DNA vectors should be cloned to stably insert this RMCE cassette into the genome of *P. pastoris* ("tagging"). Following clone evaluation, a proof of concept should be provided through a RMCE reaction in *P. pastoris*.

For the second aim, this PhD thesis focuses on the secretory production of the full ectodomain of the transmembrane serine protease mouse Tmprss2. Tmprss2 is a highly interesting drug target playing a crucial role for influenza A infectivity. Hence, the recombinant production of Tmprss2 for structural and functional analyses is an important step on the path structure based drug design. The expression of mouse Tmprss2 should be tested in a number of expression hosts, starting with *P. pastoris*. Using the most suitable host, the production and the downstream processes should be improved for the optimal production of Tmprss2. Finally, a preliminary enzymatic activity test should be performed to evaluate the ability of recombinant Tmprss2 to specifically cleave its putative viral substrate hemagglutinin H1.

2 Material and methods

2.1 Chemicals, kits and reagents

Unless stated otherwise, all chemicals used were purchased as "pro analysis" grade from Bayer, Becton Dickinson (BD), Carl Roth, GE Healthcare, Lonza, Macherey-Nagel, Merck KGaA (Merck), New England Biolabs (NEB), Qiagen, Roche, Sigma-Aldrich (Merck KGaA) and Thermo Fisher Scientific.

2.1.1 Water quality

The purified water used in this work was filtered by a 0.22 µm Milli-Q dispense system (Merck KGaA) and will be referred to as "MQ-H_2O". Sterile, purified water was obtained by autoclaving MQ-H_2O for 20 min at 121 °C. This water was treated as a substitute for bi-destilled water (ddH_2O). Unless stated otherwise all buffers and solutions were prepared with MQ-H_2O.

2.1.2 Enzymes and molecular weight standards

An overview of commonly used enzymes and molecular weight standards in this work is presented in Table 2.1 and Table 2.2.

Table 2.1. Enzymes.

Name	Supplier
Restriction Endonucleases	NEB
T4 Ligase	Roche
Antarctic Phosphatase	NEB
Phusion® Hot Start II DNA Polymerase	NEB
KOD Hot Start DNA Polymerase	Merck
RNase A	Qiagen
DNase I	SFPR, HZI

Table 2.2. Molecular weight standards.

Name	Application	Supplier
Smart Ladder	DNA agarose gel electrophoresis	Eurogentech
PageRuler Plus prestained	SDS-PAGE, Western Blot	Thermo Fisher Scientific
PageRuler unstained	SDS-PAGE	Thermo Fisher Scientific
Precision Plus AllBlue	SDS-PAGE, Western Blot	Bio-Rad
Precision Plus unstained	SDS-PAGE	Bio-Rad
Amersham HMW standard	Native PAGE	GE Healthcare
Amersham LMW standard	Native PAGE	GE Healthcare

2.1.3 Culture media & supplements

The media for yeast and bacterial cultures were autoclaved 20 min at 121 °C. Heat instable components such as glucose and antibiotics were added after sterile filtration (0.2 µm) to the autoclaved medium.

The cultivation of *E. coli* was performed in LB medium (Table 2.3). For LB agar plates, the LB medium was supplemented with 16 g/L of Bacto-Agar (BD) prior to sterilization. The regenerative cultivation phase of transformed *E. coli* was done in SOC medium (super optimal broth for catabolite repression) was used (Table 2.4). Unless noted otherwise, the appropriate antibiotics or reagents for blue-white selection were applied at the concentrations listed in Table 2.5.

Table 2.3. LB medium.

Component	Final concentration
Bacto Tryptone	10 g/L
Bacto Yeast Extract	5 g/L
NaCl	5 g/L
in MQ-H_2O	to 1 L

Table 2.4. SOC medium.

Component	Final concentration
Bacto Tryptone	20 g/L
Bacto Yeast Extract	5 g/L
KCl	186 mg/L
NaCl	50 mg/L
in MQ-H_2O	to 1 L
Supplements to add after autoclaving:	
$MgCl_2$ (1 M Stock)	10 mM
Glucose (20 % v/v Stock)	10 mM

Table 2.5. Antibiotics and supplements for bacterial cultures.

Name	Applied final concentration
Ampicillin (Amp)	100 µg/mL
Chloramphenicol (Cm)	25 µg/mL
Gentamicin (Gm)	7 µg/mL
Kanamycin (Kan)	50 µg/mL
Tetracyclin (Tet)	10 µg/mL
Zeocin (Zeo)	50 µg/mL
Bluo Gal	100 µg/mL
IPTG	40 µg/mL

Applied antibiotics and their concentrations as well as commonly used stock solutions are listed in Table 2.6 and Table 2.7. The cultivation of *P. pastoris* in suspension cultures was conducted in YPD, BMGY, BMMY, BMD or BMM media. The recipes are listed in Table 2.8 and Table 2.9. All optional supplements for different media variants are marked in italics. For agar plates 20 g/L of Bacto-Agar (BD) were added to the media prior to sterilization. Auxotrophic selection of *P. pastoris* GS115 was conducted on MD agar plates; MDH agar plates (MD supplemented with

histidine) served as comparative control plates (Table 2.10). An autoclaved solution of 1 M D-Sorbitol in MQ-H_2O was used for the cultivation phase of transformed *P. pastoris*. Cultivation in the bioreactor was performed with the media described in Table 2.11 - Table 2.15.

Table 2.6. Antibiotics for yeast cultures.

Name	Applied final concentration
Geneticin G418 (G418)	300 µg/mL
Zeocin (Zeo)	100 µg/mL

Table 2.7. Common *P. pastoris* stock solutions.

10x YNB	Final concentration
Yeast Nitrogen Base (YNB) (with ammonium sulfate, without amino acids)	134 g
in MQ-H_2O	to 1 L
→ Heated if necessary to dissolve → Filter sterilized and stored at 4 °C	
500x B	Final concentration
Biotin	20 mg
in MQ-H_2O	to 100 mL
→ Filter sterilized and stored at 4 °C	
100x H	Final concentration
L-Histidine	400 mg
in MQ-H_2O	to 100 mL
→ Filter sterilized and stored at 4 °C	
10x D	Final concentration
D-Glucose (Dextrose)	200 mg
in MQ-H_2O	to 1 L
→ Filter sterilized and stored at 4 °C	
500x GY	Final concentration
Glycerol p.a.	100 mL
in MQ-H_2O	to 1 L
→ Sterilized by autoclaving and stored at 4 °C	
10x M	Final concentration
Methanol	5 mL
in MQ-H2O	to 100 mL
→ Filter sterilized and stored at 4 °C	
1 M Potassium Phosphate Buffer pH 6.0	Final concentration
K_2HPO_4 (1 M Stock)	132 mL
KH_2PO_4 (1 M Stock)	868 mL
→ Adjusted pH if necessary with phosphoric acid or KOH → Filter sterilized and stored at room temperature	
Antifoam Struktol® J 673 A Stock	Final concentration
Struktol® J 673 A	5 % (v/v)
in methanol	to 500 mL
→ Filter sterilized and stored at room temperature	

Table 2.8. YPD(S) medium.

Component	Final concentration
Bacto Peptone	2 % (w/v)
Bacto Yeast Extract	1 % (w/v)
Only for YPDS: D-Sorbitol	1 M
in MQ-H_2O	to 900 mL
Supplements to add after autoclaving:	
10xD Stock	100 mL

Table 2.9. BMD/BMG/BMM and BMDY/BMGY/BMMY media.

Component	Final concentration
Only for BMDY/BMGY/BMMY: Bacto Yeast Extract	*1 % (w/v)*
in MQ-H_2O	to 700 mL
Supplements to add after autoclaving:	
1 M Potassium Phosphate Buffer pH 6.0 Stock	100 mL
10x YNB Stock	100 mL
500x B Stock	2 mL
Only for BMD(Y): 10x D Stock (Glucose)	*100 mL*
Only for BMG(Y): 10x GY Stock (Glycerol)	*100 mL*
Only for BMM(Y): 10x M Stock (Methanol)	*100 mL*

Table 2.10. Solid MD and MDH medium.

Component	Final concentration
Bacto Agar	7.5 g
in MQ-H_2O	to 400 mL
Supplements to add after autoclaving:	
10x YNB Stock	50 mL
500x B Stock	1 mL
Glucose (Dextrose)	50 mL
Only for MDH: 100 x H Stock	*5 mL*

Table 2.11. Bioreactor growth medium (for 2 L scale fermentation).

Component	Final concentration
Glycerol p.a.	72 g
K_2SO_4	32.8 g
$MgSO_4$x7 H_2O	26.8 g
$CaSO_4$x2 H_2O	1.67 g
in MQ-H2O → Direct autoclaving in the reactor vessel	in 1.4 L
Supplements to add after autoclaving:	
Hexametaphosphate medium	400 mL
PTM1 Trace Salt Stock	7.83 mL
500x B	7.83 mL

Table 2.12. Hexametaphosphate medium pH 5.6 (for 2 L scale fermentation).

Component	Final concentration
Sodium Hexametaphosphate	45 g
$(NH_4)_2SO_4$	16.2 g
in MQ-H2O → Adjusted pH to 5.6 if necessary with H_2SO_4 → Filter sterilized and stored at 4 °C	to 400 mL

Table 2.13. PTM1 trace salts stock (for 2 L scale fermentation).

Component	Final concentration
$CuSO_4$x5 H_2O	6.00 g
NaI	0.08 g
$MnSO_4xH_2O$	3.00 g
Na_2MoO_{4x}2 H_2O	0.20 g
$CoCl_2$	0.50 g
$ZnCl_2$	20.0 g
Fe(II)SO_4*7 H_2O	65.0 g
HBO_3	0.02 g
H_2SO_4	5.0 mL
in MQ-H_2O → Filter sterilized and stored at 4 °C	to 1 L

Table 2.14. Bioreactor glycerol feed stock (for 2 L scale fermentation).

Component	Final concentration
Glycerol p.a.	397 mL (50 % w/v)
in MQ-H2O	to 1 L
Supplements to add after autoclaving:	
PTM1 Trace Salt Stock	3.6 mL
500x B	3.6 mL

Table 2.15. Bioreactor methanol feed stock (for 2 L scale fermentation).

Component	Final concentration
Methanol	500 mL
Supplements to add after autoclaving:	
PTM1 Trace Salt Stock	6 mL
500x B	6 mL

The cell culture media used for the cultivation of insect cell lines and mammalian cell lines were purchased as ready-to-use solutions (Table 2.16). F-17 medium was supplemented with additives prior to use according to Table 2.17. An overview of all supplements is presented in Table 2.18. All required supplements were only added after filtration at 0.2 µm.

Table 2.16. Cell culture media.

Cell Culture Medium	Cell Line	Supplier
EX-CELL™ 405	Hi5	SAFC
EX-CELL™ 420	SF21	SAFC
F17	HEK293-6E	Thermo Fisher Scientific

Table 2.17. Supplements used in F17 medium.

Name	Final concentration
G418	25 µg/L
Pluronic F68	1 g/L
L-Glutamine	7.5 mM

Table 2.18. Cell culture supplements.

Name	Supplier
Fetal Calf Serum (FCS)	Thermo Fisher Scientific
G418	Thermo Fisher Scientific
Gentamicin	Thermo Fisher Scientific
Pluronic F68	Thermo Fisher Scientific
L-Glutamine	Thermo Fisher Scientific

2.1.4 Transfection reagents

Delivery of recombinant DNA to cells was performed using cationic lipofection techniques. The reagents used for each cell line and DNA construct as listed in Table 2.19.

Table 2.19. Transfection reagents.

Reagent	Application	Supplier
SuperFect	Bacmid Transfection	Qiagen
Polyethylenimine (linear, MW ~25 kDa)	HEK293-6E	Polysciences

2.2 Oligonucleotides and plasmids

Oligonucleotides were used for sequencing, cDNA synthesis and PCR (including fusion PCR and RT-PCR). All oligonucleotides in this work have either been obtained from specific kits or were ordered at Eurofins Genomics (Ebersberg, Germany) in HPSF or HPLC purified quality. For simplification, DNA oligonucleotides that were used for PCR or for DNA sequencing will be referred to as "primers" further on, unless specifically stated otherwise. Thereby, the primer binding to the defined 5' region of the template DNA will be designated as the forward primer, while the primer binding complementary to the respective 3' region will be referred to as the reverse primer. A complete list of all used primers can be found in Appendix I. The principal plasmids used in this work are listed and explained in Table 2.20. An overview of the vectors that were generated in this work is listed below (§ 2.3).

Table 2.20. Plasmids.

***P. pastoris* Expression vectors**
pPICZ(α) / pGAPZ(α) (Thermo Fisher Scientific) Vector family for protein expression using the *AOX1* promoter (pPICZ) or the *GAP* promoter (pGAPZ). The letter "α" indicates the presence of a *S. cerevisiae* α-mating factor secretion signal (MF-α ss) on the vector for secretory expression. The plasmids harbor a Zeocin resistance cassette for dual selection in *E. coli* and in *P. pastoris*. Therein, the Zeocin resistance gene *Sh ble* is flanked by a TEF1/EM7 fusion promoter and a CYC1 terminator. **pPIC9K (Thermo Fisher Scientific)** Expression vector which harbors the *AOX1* promoter. It also includes the 3'*AOX1* genomic region to mediate gene replacement upon transformation. This vector is usually applied for the generation of mut^S strains by replacing the *AOX1* gene with the GOI and the remaining vector sequence. Selection in *E. coli* is performed with ampicillin. For auxotrophic selection in *P. pastoris*, pPIC9k harbors a *his4* gene. Furthermore, a bacterial Tn903 kanamycin resistance gene is present on the vector, which is used to screen for *P. pastoris* multi copy insertion clones against an increasing gradient of G418. **pBGP1 (Courtesy of David Resina, Bioingenium S.L., reference Lee et al. (2005))** Episomal expression vector is based on pGAPZα-A (Thermo Fisher Scientific). A PARS1 sequence was added to the backbone of pGAPZα-A in order to facilitate episomal replication of the plasmid in *P. pastoris*. Furthermore, the pUC origin of replication (ori) was replaced by a combination of the pBR322 ori and a bacterial ampicillin resistance cassette to allow for a more cost-efficient bacterial selection.
***E. coli* Vectors**
pJET1.2 (Thermo Fisher Scientific) Intermediate vector for blunt cloning of PCR products. The vector was obtained as part of the CloneJET PCR Cloning Kit. **pET22B(+) (Merck)** Expression vector for *E. coli*, which harbors a pelB signal for periplasmic localization of the recombinant target protein.
Vectors used in cell culture
pFlpBtM-III (Steffen Meyer, HZI) Third version of the versatile multi-host expression. It is applicable for transient expression in HEK293-6E cells and for the creation of bacmids using TN7 based transposition. Furthermore, it serves as a RMCE donor vector for CHO Lec 3.2.8.1 cell lines employing FRT F_3 and F_{WT} sites. For secretory expression, pFlpBtM-III offers an IgG secretion signal. **pTTo/GFPq (NRC, BRI, Montreal, Canada)** A control vector that is used to monitor the transfection efficiency in HEK293-6E cells. It harbors a GFPq gene (red shifted green fluorescent protein) for detection via flow cytometry.

Parental Construction Vectors
FPB-31-902-RudolphRFP (ATUM) Commercially obtained vector carrying the gene for the red fluorescent marker RudolphRFP. **pBGP1 (Courtesy of David Resina, Bioingenium S.L., reference Lee et al. (2005))** The fragment comprising the ampicillin resistance cassette and the pBR322 ori was taken from this vector to create the vector pGAPZAΔ8 from pGAPZα-A in this work. Furthermore, it served as a basis to generate the Flp recombinase expressing RMCE helper plasmid for *P. pastoris*. **pFlpBtM-III-mCherry (Steffen Meyer, HZI)** The gene of the fluorescent marker mCherry was cloned from this vector. Furthermore, the human PGK (hPGK) promoter was cloned from this vector for the selection trap analysis in *P. pastoris*. **pGAPZα-A (Thermo Fisher scientific) and pGAPZAΔ8** This vector was used to clone the vector pGAPZAΔ8, which served as a basis to generate the *P. pastoris* RMCE vector pYTA and its variants as well as the exchange vector pYEX. In comparison to pGAPZα-A, pGAPZAΔ8 harbors the ampicillin resistance cassette and the pBR322 ori of pBGP1 (see above). The MF-α ss was removed from this vector. The vector product pGAPZAΔ8 misses 8 bp at the original position of the MFα-ss compared to the commercially available pGAPZ-A. This is reflected in the suffix "Δ8". **pPIC9K (Thermo Fisher scientific)** This vector was used as a source to clone the genomic 3'*AOX1* region and the *his4* gene into respective *P. pastoris* RMCE tagging vectors. **pOG44 (Thermo Fisher Scientific)** Flp recombinase expression vector of the Flp-In™ system (Thermo Fisher Scientific). The flp gene was amplified from this plasmid to generate the RMCE helper plasmid for *P. pastoris*. **pOpIE1-eGFP-HA (Maren Bleckmann, HZI)** The OpIE1 promoter was taken from this vector to be tested as a possible alternative to the hPGK promoter on pFlpBtM for the multi-host compatible induction of the selection trap upon RMCE. **pUC57-A1:hsABCB6-SyntheticSegments (Christian Kambach, HZI)** The y-eGFP gene was amplified from this plasmid to be used as model GOI in *P. pastoris*. **pUC57-SplGFP (Synthesis by GenScript, Hong Kong, China)** Commercially obtained vector that harbors two synthetic fragments used to generate the *P. pastoris* RMCE tagging vector. The first fragment comprises "BstBI – FRT F3 – MCS – NotI", which includes a modified multi cloning site (MCS). The second fragment consists of PciI – ATG – FRT F_{WT} – HindIII – NcoI". "ATG" denotes a start codon to complement the RMCE selection trap. **pUC57-mTMPRSS2-D343N-FLAG-8xHis (Synthesis by GenScript, Hong Kong, China)** Commercially obtained vector carrying the gene of the *P. pastoris* optimized ectodomain of mouse Tmprss2-D343N with a C-terminal combination of TEV protease site, 1xFlag tag and 8xHis tag.

2.3 Generated DNA vectors in this work

A number of DNA plasmids was designed and cloned for the projects of this PhD thesis. In this section, the generation of these vectors, including the cloning strategy, is described in detail.

pGAPZAΔ8, pYEXs and pYTA

The RMCE tagging vector for *P. pastoris* was cloned from the plasmid pGAPZα-A (Thermo Fisher Scientific). First, the MF-α ss sequence was deleted through digestion with *Bst*BI and *Eco*RI. The

sticky ends were filled in and the vector was blunt ligated using the CloneJET PCR Cloning Kit. Furthermore, a fragment comprising an ampicillin resistance cassette and a pBR322 origin of replication (ori) from the vector pBGP1 (§ 2.2) was cloned into the plasmid, replacing the pUC ori, through digestion with *Pci*I and *Bgl*II. The vector product was designated as pGAPZAΔ8. The term "Δ8" indicates that the vector is missing 8 bp between the *Bst*BI and *Eco*RI restriction sites compared to the commercially available pGAPZ-A (Thermo Fisher Scientific).

The FRT sites (F_3, F_{WT}) and the Δ*his4* gene were cloned into pGAPZAΔ8 using two synthetic DNA modules (GenScript). The first synthetic module comprised a 5'-*Bst*BI site, FRT F_3, a modified MCS and a 3'-*Not*I site (Synth-A). The second module comprised a 5'-*Pci*I site, a *Pst*I site, an ATG codon, an *Nde*I site, FRT F_{wt}, a *Hin*dIII site and a 3'*Nco*I site (Synth-B). The *his4* gene was PCR amplified in two parts from pPIC9k (Thermo Fisher Scientific) without its ATG start codon (Δ*his4*). The parts were defined by its native *Nco*I site at 880 bp of the gene. The 5'-fragment of Δ*his4* (Δhis4-A; 4-880 bp) was amplified with the primers HindIII-dHis4-F and dHis4-880NcoI-R. Hence, the product displayed 5'-*Hin*dIII and 3'-*Nco*I sites. The 3'-fragment (Δhis4-B; 880 bp-2532 bp), including the downstream transcriptional terminator sequence, was amplified with the primers 880NcoI-dHis4-F and dHis4T-ScaI-R, adding a 3'-*Sca*I site to it in addition to the native 5'-*Nco*I. The vector Δhis4-B was digested with *Nco*I and *Sca*I and cloned into pGAPZAΔ8, which was digested with *Nco*I and *Eco*RV. Thereby, Δhis4-B was placed downstream of the TEF1/EM7 promoter, replacing the Zeocin resistance gene (*Sh ble*) and part of its CYC1 terminator. The validated vector product was digested between the TEF1/EM7 promoter and Δhis4-B by *Nco*I. Synth-B was inserted at this site after digestion with *Pci*I and *Nco*I. Δhis4-A was inserted downstream of FRT F_{WT} via *Hin*dIII and *Nco*I, completing the Δ*his4* CDS. The vector product was designated as pGAPZAΔ8-Fwt-his4+, as it featured a complemented Δ*his4* gene under the control of the TEF1/EM7 promoter. Next, Synth-A was inserted into pGAPZAΔ8-Fwt-his4+ via *Bst*BI and *Not*I, replacing its original MCS with the "FRT F_3 – modified MCS" combination. The resulting vector was named pYEXs ("yeast exchange simulation").

The *y-eGFP* gene was cloned into pYEXs as a model GOI. For this, the gene was PCR amplified from pUC57-A1:hsABCB6-SyntheticSegments (Christian Kambach, HZI) with the primers EcoRI-yeGFP-F and SacII-yeGFP-R and cloned into the MCS of pYEXs using *Eco*RI and *Sac*II. The vector product was named pYEXs-yeGFP. The human PGK1 (hPGK) promoter from pFlpBtM-III (Steffen Meyer, HZI) and the OpIE1 promoter from *Orgyia pseudotsugata* from pOpIE1-eGFP-HA (Maren Bleckmann, HZI) were respectively cloned into pYEXs-yeGFP, replacing the TEF1/EM7 promoter upstream of Δ*his4*. For this, the hPGK promoter was PCR amplified with the primers BamHI-PGK-F and PGK-PstI-R. The OpIE1 promoter was PCR amplified with the primers BamHI-OpIE1-F and PstI-OpIE1-R. The PCR products were cloned into pYEXs-yeGFP through *Bam*HI and *Pst*I. The resulting plasmids were named pYEXsPGK-yeGFP and pYEXsOpi-yeGFP.

The tagging vector named pYTA ("yeast tagging") was generated from pYEXs. The Zeocin resistance cassette ("TEF1/EM7 promoter – *Sh ble* – CYC1 terminator") was amplified by PCR from pGAPZAΔ8 with the primers BamHI-TEF1-F and Cyc1T-NdeI-R, adding a 5'-*Bam*HI and a 3'-*Nde*I site. The PCR product was cloned into pYEXs through *Bam*HI and *Nde*I, replacing the "TEF1-EM7 promoter – ATG start codon" combination. The *RudolphRFP* gene was incorporated into pYTA by cloning it directly from the plasmid FPB-31-902-RudolphRFP (ATUM) into the MCS of pYTA via *Eco*RI and *Not*I. The resulting vector was designated as pYTA-RudolphRFP.

Fluorochrome test vectors (pPICZ / pGAPZAΔ8 - RudolphRFP / mCherry / y-eGFP)

To generate the *GAP* and *AOX1* test cell lines based on pGAPZAΔ8 and pPICZ-A (Thermo Fisher Scientific), the genes of RudolphRFP, y-eGFP and mCherry were cloned into the MCS of the two vectors through *Eco*RI and *Not*I. *RudolphRFP* and *y-eGFP* were directly cloned from pYTA-RudolphRFP and pYEXs-yeGFP. The *mCherry* gene was PCR amplified from the vector pFlpBtM-II-mCherry (Steffen Meyer, HZI) with the primers EcoRI-mCherry-F and NotI-mCh(woTAG)-R for cloning. The six vector products were named: pPICZ-RudolphRFP, pPICZ-yeGFP, pPICZ-mCherry, pGAPZAΔ8-RudolphRFP, pGAPZAΔ8-y-eGFP and pGAPZAΔ8-mCherry.

pYTAaox-mCherry and pYTAaox-GAP-mCherry

For the generation of the *AOX1* tagging vector pYTAaox-mCherry, the vector pYTA-RudolphRFP was modified. The *AOX1* promoter was cloned directly into pYTA from pPICZ-A (Thermo Fisher Scientific) using *Bgl*II and *Bst*BI. The 3'-*AOX1* region was cloned into the vector via *Pci*I, following its PCR amplification from pPIC9k (Thermo Fisher Scientific) with the primers PciI-3'AOX-f and PciI-3'AOX-r. The resulting vector was designated as pYTAaox. The *mCherry* gene was cloned into pYTAaox through *Eco*RI and *Not*I following PCR amplification from the vector pFlpBtM-II-mCherry (Steffen Meyer, HZI) with the primers EcoRI-mCherry-F and NotI-mCh(woTAG)-R. The resulting tagging vector was named pYTAaox-mCherry. The *GAP* promoter was inserted into the RMCE cassette of pYTAaox-mCherry downstream of FRT F_3 in an additional cloning step. The *GAP* promoter was PCR amplified from pGAPZAΔ8 with the primers BsaI-MfeI-GAP-F and EcoRI-GAP-R, thereby adding a 5'-BsaI (MfeI) site and a 3'-EcoRI site to the fragment. The PCR product was digested with *Bsa*I and *Eco*RI and ligated into pYTAaox-mCherry (digested with *Eco*RI). The vector product was named pYTAaox-GAP-mCherry.

Variants of the RMCE tagging vector pYTA to relocate FRT F_3

In total, four additional variants were generated from the first version of the tagging vector pYTA-RudolphRFP (V1). For variant V2 (pYTAΔF3-RudolphRFP), a fragment comprising FRT F_3 and *RudolphRFP* was isolated from pYTA-RudolphRFP digested with *Bst*BI and *Not*I and replaced by the corresponding region of the vector pGAPZAΔ8-RudolphRFP. The vector product was named

pYTAΔF3-RudolphRFP, since this cloning step effectively resulted in the deletion of F_3 from the tagging vector (ΔF_3).

In variant V3 (pYTAF3PR-RudolphRFP), FRT F_3 was moved into the promoter region (PR) of the *GAP* promoter, upstream of its TATA box. FRT F_3 was inserted into the *GAP* promoter by fusion PCR (see chapter 2.6.1). Two fragments comprising vector regions upstream and downstream of the TATA box of the *GAP* promoter were amplified by PCR from pYTA-RudolphRFP (V1). The 5'-fragment (fragment A) comprised the *GAP* promoter region upstream of the TATA box, the TATA box itself and a partial sequence of FRT F_3. The 3'-fragment (fragment B) consisted of the remaining sequence of FRT F_3 (overlapping with the sequence on fragment A), the TATA box, the continuing *GAP* promoter region and the gene encoding *RudolphRFP*. The fusion PCR was conducted with FRT F_3 serving as the overlapping fusion sequence. The final PCR product comprised the modified *GAP* promoter and *RudolphRFP* with 5'-*Spe*I and 3'*Not*I sites. The used primers are listed in Table 2.21. The PCR product was finally used to replace the *GAP* promoter of pYTA-RudolphRFP through cloning with *Spe*I and *Not*I.

Table 2.21. Primers of the fusion PCR to insert FRT F_3 into the *GAP* promoter.
FRT F_3 added in two parts in two rounds of PCR to a location upstream of the TATA box of the *GAP* promoter. The sequences of the oligonucleotides are listed in the appendix. Italics indicate the internal primers that were used to introduce the modification.

Fragment	Product	Forward primer	Reverse primer
A	*Spe*I-(5'PGAP)-$F_{3\text{-part1}}$	SpeI-PGAP-F	*GAPintF3-R*
B	$F_{3\text{-part2}}$-(3'PGAP)-RudolphRFP-*Not*I	*GAPintF3-F*	NotI-Rudolph-R
A+B (Fusion)	*Spe*I-PGAP-F_3PR-RudolphRFP-*Not*I	SpeI-PGAP-F	NotI-Rudolph-R

To generate variant V4 (pYTAatgF3-RudolphRFP), the *GAP* promoter was amplified by PCR from pYTA-RudolphRFP with the oligonucleotide primers SpeI-PGAP-F and BstBI-PGAP-ATG-R, which added a 5'-*Spe*I site, the modified sequence and a 3'-*Bst*BI to the PCR product. The PCR product was cloned into pYTA-RudolphRFP through *Spe*I and *Bst*BI, replacing the original *GAP* promoter with the modified version. The vector product harbored the ATG codon upstream of F_3.

In variant V5 (pYTAUR-RudolphRFP), the *GAP* promoter was placed downstream of FRT F_3 inside the RMCE cassette of pYTA-RudolphRFP. To enable the integration in the *GAP* locus by homologous recombination, a 508 bp chromosomal upstream region of *GAP* (URG) from *P. pastoris* GS115 (Genbank #FN392320, position 809,079 – 809,598) was added to the vector upstream of FRT F_3. The URG was amplified by PCR from isolated genomic DNA of *P. pastoris* GS115 (§ 2.6.1, 2.6.11) and a combination of "5'-*Xba*I – FRT F_3 – *Spe*I – 3'-*Not*I" was added to its 3'-end for cloning. This was done in three rounds of PCR with the XbaI-URG-F serving as the forward primer. The primers URG-F3-R1, SpeI-URG-F3-R2 and Not-Spe-F3_R served as the respective reverse primers. The resulting "URG – FRT F_3" product was digested with *Xba*I and *Not*I and ligated into pYTA-RudolphRFP (digested with *Spe*I and *Not*I), removing the "*GAP*

promoter – F_3 – *RudolphRFP*" sequence. The "*GAP* promoter –*RudolphRFP*" combination from pGAPZAΔ8-RudolphRFP was subsequently cloned into the vector via *Spe*I and *Not*I.

pYEXsUR-yeGFP, pYEXUR-yeGFP and pBGP1KΔ8-FlpWT

The *P. pastoris* exclusive RMCE donor vector was generated based on pYTAUR-RudolphRFP (V5). First, *RudolphRFP* and the Zeocin resistance cassette of pYTAUR-RudolphRFP were replaced with a fragment from pYEXs-yeGFP comprising the *y-eGFP* gene and the "TEF1 promoter – ATG start codon" combination using *Eco*RI and *Nde*I. The resulting vector with the complemented selection trap Δ*his4* was named pYEXsUR-yeGFP. To generate the RMCE donor vector pYEXUR-yeGFP, Δ*his4* was deleted from this vector through digestion with *Hin*dIII and *Eco*RV. The sticky ends were filled in and vector was blunt ligated using the components of the CloneJET PCR Cloning Kit.

The Flp recombinase delivering RMCE helper vector pBGP1KΔ8-FlpWT was designed based on the episomal PARS1 plasmid pBGP1 (Lee et al. 2005). The *Sh ble* gene on pBGP1 was exchanged for the *Tn903* kanamycin resistance gene as a feasible alternative for direct selection on G418 (Lin-Cereghino et al. 2008, Papakonstantinou et al. 2009). The *Tn903* gene was amplified from pPIC9k (Thermo Fisher Scientific) in two rounds of PCR with the forward primer NcoI-kan_f and the respective reverse primers kan_r1 and ScaI-kan_r2. The PCR product comprised a 5'-*Nco*I site, the *Tn903* gene, a *Pst*I site, the first 27 bp of the CYC1 terminator and a 3'-*Sca*I site. It was digested with *Nco*I and *Sca*I and ligated into pBGP1 (digested with *Nco*I and *Eco*RV). The resulting vector was designated as pBGP1K. The *GAP* promoter, the MF-α ss and the MCS of pBGP1K were replaced by the corresponding fragment from pGAPZAΔ8 using *Spe*I and *Bam*HI. Thereby, the MF-α ss was effectively deleted from pBGPK1. The vector product was designated as pBGP1KΔ8. The *FLP* gene was taken from pOG44 (Thermo Fisher Scientific) and amplified by PCR with the oligonucleotide primers Bsa-Flpwt-F and Not-Flpwt-R, adding a 5'-*Bsa*I (*Eco*RI) site and a 3'-*Not*I site to its ends. *Bsa*I was used to create an *Eco*RI overhang. The PCR product was cloned into the MCS of pBGP1KΔ8 using *Eco*RI and *Not*I, resulting in the vector pBGP1KΔ8-FlpWT.

Tmprss2 expression vectors based on pPICZα-A and pFlpBtM-III / pFlpBtM-III-insect

The *Tmprss2* genes were amplified in two rounds of PCR with the forward primer BsaI-TMP-F and the reverse primers TMP-R1 and NotI-TMP-R2, respectively, to add the CDS of the C-terminal tags ("TEV site – 1xFlag tag – 8xHis tag"). Thereby, a 5'-*Bsa*I (*Eco*RI) site and a 3'-*Not*I site were added for cloning. The PCR products were digested accordingly and cloned into pPICZα-A (digested with *Eco*RI and *Not*I). The resulting vectors were designated as pPICZα-Tmprss2-WT and pPICZα-Tmprss2-D343N.

The vectors to express the serine protease domain (SPD) of Tmprss2 (amino acids 254 – 490; gene positions 760 – 1470 bp of the native, full length protein) in *P. pastoris* were generated based on pPICZα-A. The SPD of *Tmprss2-WT* and *Tmprss2-D343N* including the tags was respectively amplified by PCR from pPICZα-Tmprss2-WT and pPICZα-Tmprss2-D343N with the primers BsaI-TMPSPD-F and NotI-TMP-R2, which added 5'-*Bsa*I (to create an *Eco*RI overhang) and 3'-*Not*I sites. The PCR products were subsequently cloned into pPICZα-A with *Bsa*I / *Eco*RI and *Not*I. The vectors were designated as pPICZα-Tmprss2-SPD-WT and pPICZα-Tmprss2-SPD-D343N.

The generation of expression vectors for *P. pastoris* with codon optimized constructs was also based on pPICZα-A. The *Tmprss2-D343N* gene was codon optimized for *P. pastoris* by GenScript (*Tmprss2-D343Nopt*), including the tags ("TEV site – 1xFlag tag – 8xHis tag"). The native *Eco*RI of *Tmprss2* was removed during the gene optimization. The synthetic fragment was flanked by a 5'-*Eco*RI site and a 3'-*Not*I site. The wildtype gene *Tmprss2-WTopt* was obtained by fusion PCR from *Tmprss2-D343Nopt* as described in chapter 2.6.1. An overview including the used primers is listed in Table 2.22. *Tmprss2-WTopt* and *Tmprss2-D343Nopt* were cloned into pPICZα-A via *Eco*RI and a *Not*I. The vectors were named pPICZα-A-Tmprss2-WTopt and pPICZα-A-Tmprss2-D343Nopt.

Table 2.22. Primers of the fusion PCR to generate *Tmprss2-WTopt*.
The gene *Tmprss2-WTopt* was generated by fusion PCR to a location upstream of the TATA box of the *GAP* promoter. The sequences of the oligonucleotides are listed in the appendix. Italics indicate the internal primers that were used to introduce the modification.

Fragment	Product	Forward primer	Reverse primer
A	*Eco*RI-tmprss2-WTopt$_{part1}$	EcoRI-Tmprss2-opt-F	*int-TMPopt-mut-R*
B	tmprss2-WTopt$_{part2}$-*Not*I	*int-TMPopt-mut-F*	NotI-Tmprss2-opt-R
A+B (Fusion)	*Eco*RI-tmprss2-WTopt-*Not*I	EcoRI-Tmprss2-opt-F	NotI-Tmprss2-opt-R

The vector pPICZ-HSA was generated based on pPICZα-A (Thermo Fisher Scientific). The *HSA* gene was amplified by PCR from isolated genomic DNA of the *P. pastoris* GS115 HSA mut^S control strain that is available as part of the Pichia Expression Kit (Thermo Fisher Scientific). The recombinant *HSA* gene in this strain comprises the natural CDS, except for a silent point mutation to remove a native *Sac*I site. The PCR was conducted with the primers BstBI-HSA-FOR and XhoI-HSA-MCS-REV to add 5'-*Bst*BI and 3'-*Xho*I sites to the product for cloning as well as a slightly modified MCS fragment for the vector, which does not include the *Sfi*I site anymore in comparison to the original MCS, as this enzyme cuts inside *HSA*. The "HSA-MCS" PCR construct was cloned into pPICZα-A via *Bst*BI and *Xho*I. The vector was designated as pPICZ-HSA. The "HSA-MCS" sequence forms a common ORF with the inherent c-myc and 6xHis tags of the vector to also enable stand-alone expression of tagged HSA. The codon optimized gene *Tmprss2-D343Nopt* was cloned

directly into pPICZ-HSA from pPICZα-Tmprss2-D343Nopt by *Eco*RI and *Not*I. The resulting vector was designated as pPICZ-HSA-Tmprss2-D343Nopt.

The multi-host vector pFlp-BtM-III (Steffen Meyer, HZI) was used for the expression of Tmprss2 in HEK2936E cells. For the expression in SF21 and Hi5 insect cell lines, a newly developed derivative of pFlpBtM-III, pFlpBtM-III-insect (Maren Bleckmann, HZI), was used. The genes of *Tmprss2-WT* and *Tmprss2-D343N* including the tags were amplified by PCR from pPICZα-Tmprss2-WT and pPICZα-Tmprss2-D343N with the primers NheI-TMP-F and NotI-TMP-R2. The PCR fragments were cloned in frame with the IgG ss into pFlpBtM-III and pFlpBtM-III-insect using *Nhe*I and *Not*I. The vector products were designated as pFlpBtM-III-Tmprss2-WT, pFlpBtM-III-Tmprss2-D343N, pFlpBtM-III-insect-Tmprss2-WT and pFlpBtM-III-insect-Tmprss2-D343N.

2.4 Strains and cell lines

All cloning works were performed in *E. coli* Top10. The generation of recombinant bacmids was performed in *E. coli* DH10 strains harboring the EmBacY bacmid and a transposase expressing helper plasmid (DH10EmBacY). All *E. coli* strains are listed in Table 2.23. The RMCE system was established in *P. pastoris* strain GS115. The expression of Tmprss2 was performed in strain KM71H. The strains are listed in Table 2.24. Table 2.25 depicts the cell lines that were evaluated as expression hosts for Tmprss2.

Table 2.23. Bacterial strains.

E. coli strain	Genotype	Supplier
Top10	F- *mcrA Δ(mrr-hsdRMS-mcrBC) φ80 (lacZ) ΔM15 ΔlacX74 recA1 araD139 Δ(ara, leu) 7697 galU galK rpsL (Str^R) endA1 nupG λ-*	Thermo Fisher Scientific
DH10(EmBacY)	F- *mcrA Δ(mrr-hsdRMS-mcrBC) Φ80(lacZ) ΔM15ΔlacX74 recA1 endA1 araD139 Δ(ara, leu)7697 galU galK l- rpsL nupG /BacloxP/ pBADZ-His6Cre/pMON7124*	EMBL (Berger)

Table 2.24. Yeast strains.

E. coli strain	Genotype	Methanol utilization	Supplier
GS115	*his4*	mut$^+$	Thermo Fisher Scientific
KM71H	*aox1::ARG4, arg4*	mut^S	Thermo Fisher Scientific

Table 2.25. Cell lines.

Cell Line	Organism of origin	Reference
HEK293-6E	Human embryonic kidney epithelium	Durocher et al. (2002)
IPLB-SF-21 AE (Sf21)	Ovaries of *Spodoptera frugiperda*	Vaughn et al. (1977)
BTI-Tn-5B1-4 (Hi5)	Ovaries of *Trichoplusia ni*	Wickham and Nemerow (1993)

2.5 Software

Microsoft Word was used for text editing. Graphic design and image editing was performed with Microsoft PowerPoint, Adobe Photoshop and GIMP. Table-based batch calculations and generation of graphs were performed in Microsoft Excel. VectorNTI suite 8.0 (Thermo Fisher Scientific) was used for PCR primer design, the development of cloning strategies and to perform *in silico* restriction analyses and protein translation. For *in silico* genome browsing for cloning and analyses, the sequencing data of *P. pastoris* GS115, published by De Schutter et al. (2009) was used. The data was obtained from the NCBI/GenBank (Bethesda, Maryland, USA) with the accession numbers FN392319, FN392320, FN392321 and FN392322 for *P. pastoris* chromosomes 1-4, respectively. DNA and protein sequence alignments were conducted with MultAlin (Corpet 1988). To aid the design of truncated versions of Tmprss2, predictions were made with Phyre2 (Kelley et al. 2015). For RNA secondary structure predictions, the web-tool RNAFold of the Vienna RNA Websuite was used (Gruber et al. 2008).

2.6 Molecular biological methods

The protocols for the molecular biological methods in this work have been adapted from standard collections (Sambrook 2001). All prepared DNA plasmids were subjected to validation by restriction digestion (§ 2.6.3) and sequencing. Sequencing was performed at the at the Genome Analytics platform GMAK at the HZI Braunschweig (Germany).

2.6.1 Polymerase chain reaction (PCR)

PCR was used to amplify target DNA from plasmid or genomic DNA templates either for cloning or for analysis. For cloning reactions, the PCR was conducted with the Phusion® Hot Start II DNA Polymerase (NEB). The standard mix and the default program are listed in Table 2.26 and Table 2.27. For the amplification of genomic DNA, the standard PCR mix from Table 2.26 was slightly modified. Instead of the 5x HF-buffer the provided 5x GC-buffer was used and the PCR mix was additionally supplemented with 5 % (v/v) DMSO. The PCR program (Table 2.27) was also slightly modified by doubling the cycle times (C1: 1 min, C2: 1 min and C3: 1 min/1 kb).

Table 2.26. Standard PCR mix with Phusion® Hot Start II DNA Polymerase.

Component	Concentration
5x HF Reaction buffer	20 µL
Forward primer (10 µM)	2.5 µL
Reverse primer (10 µM)	2.5 µL
dNTP mix (10 mM)	2 µL
Template DNA	~15 ng
Phusion Hot Start II High-Fidelity DNA Polymerase (2 U/µL)	1 µL
MQ-H_2O	x
Total volume	100 µL

Table 2.27. Default PCR Program with Phusion® Hot Start II DNA Polymerase.

Step	Temperature	Time	Cycle Repeats (C1-C3)
Initial denaturation	98 °C	5:00 min	
C1: Denaturation	98 °C	0:30 min	
C2: Annealing (adjusted to fit PCR primer Tm)	52-72 °C	0:30 min	30x
C3: Extension	72 °C	0:30 min / 1 kb	
Final Extension	72 °C	10 min	
Pause	16 °C	∞	

Fusion PCR was performed to introduce internal mutation or new sequences to a template DNA. Briefly, two fragments are initially amplified from the desired template DNA. The first fragment (A) covers the 5' region up to the position at which the mutation should be introduced. The second fragment (B) equally covers the 3' region. Both fragments are designed to harbor overlapping ends of 15 bp – 20 bp with the respective other fragment. The fusion of both fragments is performed in another PCR comprising two phases that are performed in direct succession. In the first phase, both fragments are fused via their overlapping borders and the free 3' overhangs are filled up by the polymerase. In the second phase, forward and reverse primers for the whole construct are added and the fused DNA is subsequently amplified. In this thesis, the fusion PCRs were carried out with the Phusion® Hot Start II DNA Polymerase (NEB). The PCR to amplify the fragments A and B was performed with the standard mix and the default program (Table 2.26, Table 2.27). The reaction mix and the PCR program for the fusion PCR, are listed in Table 2.28 and Table 2.29.

Table 2.28. PCR mixes for fusion PCR with Phusion® Hot Start II DNA Polymerase.

Mix I – Components:	Concentration
5x HF Reaction buffer	12 µL
Forward primer (10 µM)	2.5 µL
Reverse primer (10 µM)	2.5 µL
dNTP mix (10 mM)	3 µL
DNA Fragment A	~30 ng
DNA Fragment B	~30 ng
Phusion Hot Start II High-Fidelity DNA Polymerase (2 U/µL)	1.2 µL
MQ-H_2O	x
Total volume	60 µL (3 x 20 µL)
Mix II – Components:	Concentration
5x HF Reaction buffer	5 µL
Forward primer (10 µM)	2.5 µL
Reverse primer (10 µM)	2.5 µL
Phusion Hot Start II High-Fidelity DNA Polymerase (2 U/µL)	0.5 µL
MQ-H_2O	11.5 µL
Total volume	20 µL

Table 2.29. PCR Program for fusion PCR with Phusion® Hot Start II DNA Polymerase.

Phase I – Steps:	Temperature	Time	Cycle Repeats (C1-C3)
Initial denaturation	98 °C	10:00 min	
C1: Denaturation	98 °C	5:00 min	
C2: Annealing (adjusted to fit PCR primer Tm)	52-72 °C	1:30 min	7x
C3: Extension	72 °C	1:00 min / 1 kb	
Final Extension	72 °C	∞	
Addition of Mix II			
Add 5 µL of reaction mix II per 20 µL of reaction mix I			
Phase II – Steps:	Temperature	Time	Cycle Repeats (C1-C3)
Initial denaturation	98 °C	5:00 min	
C1: Denaturation	98 °C	0:30 min	
C2: Annealing (adjusted to fit PCR primer Tm)	52-72 °C	0:30 min	30x
C3: Extension	72 °C	0:30 min / 1 kb	
Final Extension	72 °C	10 min	
Pause	16 °C	∞	

For the analysis of bacmid DNA, the KOD Hot Start DNA Polymerase was used. The standard PCR mixture and the PCR program are listed below (Table 2.30, Table 2.31). All obtained PCR products were analyzed by agarose gel electrophoresis (see 2.6.2). If necessary, the PCR products were subsequently purified by gel extraction (§ 2.6.7.1).

Table 2.30. Standard PCR mix with KOD Hot Start DNA Polymerase.

Component	Concentration
10x KOD Reaction buffer	5 µL
$MgSO_4$ (25 mM)	2 µL
Forward primer (10 µM)	1.5 µL
Reverse primer (10 µM)	1.5 µL
dNTP mix (10 mM)	1 µL
Template DNA	1 µL
KOD Hot Start DNA Polymerase (1 U/µL)	1 µL
MQ-H_2O	37 µL
Total volume	50 µL

Table 2.31. Default PCR Program with KOD Hot Start DNA Polymerase.

Step	Temperature	Time	Cycle Repeats (C1-C3)
Initial denaturation	94 °C	2:00 min	
C1: Denaturation	94 °C	0:15 min	
C2: Annealing (adjusted to fit PCR primer Tm)	60 °C	0:30 min	30x
C3: Extension	72 °C	0:20 min / 1 kb	
Final Extension	72 °C	10 min	
Pause	16 °C	∞	

2.6.2 Agarose gel electrophoresis

DNA agarose gel electrophoresis was performed with gels comprising 1 % (w/v) agarose in 1x TAE buffer (Table 2.32). Prior to use, the agarose was completely dissolved by heating. After cooling to ~60 °C the mixture was cast on prepared gel trays and supplemented with

ethidium bromide (10 mg/mL stock solution) or Roti®-GelStain (Carl Roth) at a dilution 1.5 µL staining substance per 100 mL gel. For loading into the sample pockets, the DNA samples were mixed with a 6x DNA loading buffer (Table 2.32). Gels were run in 1x TAE buffer at 10 to 15 V/cm for 30 min and documented under UV illumination at 254 nm. 1 kb SmartLadder (Eurogentec) was used as a molecular weight standard in all runs (Table 2.2).

Table 2.32. Buffers for DNA electrophoresis.

Buffer	Composition
TAE	35.4 mM Trizma® base 0.1 % (v/v) glacial acetic acid 1 mM EDTA pH 8.5
6x DNA Loading Buffer	0.4 % (w/v) glycerol p.a. 10 % (v/v) 100x bromphenol blue 0.03 M EDTA

2.6.3 Restriction digestion of DNA

All restriction digestions were performed with enzymes by NEB according to the manufacturer's protocols with 0.5 U/µL of each restriction enzyme. If possible, the High Fidelity (HF) versions of the enzymes were used. Analytical restriction digests were conducted at 10 µL scale with 1.5 µg to 3 µg of plasmid DNA and a reaction time of 60 min. Preparative restriction digests were performed at 100 µL scale with 5 µg to 10 µg of plasmid DNA and a reaction time of 120 min to 180 min.

2.6.4 Ligation of DNA fragments

The integration of digested DNA fragments (inserts) into a compatibly digested vector backbone was performed by T4 ligation. All digested vector backbones were dephosphorylated for 15 min at 37 °C with Antarctic phosphatase (NEB) according to the supplier's protocol to avoid religation. The Antarctic phosphatase was inactivated by incubating the reaction mix at 70 °C for 15 min. The ligation reaction was conducted in 20 µL volume with 1 U T4 DNA ligase (Roche) according to the supplier's manual. The smaller fragment, usually the insert, was applied in a 4:1 molar excess over the vector backbone. Two controls were set up for each reaction. In the first control, no insert was added to the mix (religation control). In the second control, neither insert nor T4 ligase were added (negative control). The ligation was performed for 60 min to 120 min at room temperature.

The blunt ligation of PCR products or DNA fragments into the vector pJET1.2 for amplification and sequencing was performed with the CloneJET PCR Cloning Kit (Thermo Fisher Scientific) according to the manufacturer's manual with the following modifications. All reactions were performed according to the "Sticky End Protocol". About 3 - 5 µL of PCR product were used for the reaction. The blunting reaction was extended to 15 min at 70 °C. The ligation time was also extended to 30 min at room temperature. The components of the kit were also used to blunt and

religate digested DNA vector backbones following the same protocol. In this case the digested vector backbone was handled like a PCR insert and pJET1.2 was not added to the reaction mix.

2.6.5 Preparation of competent *E. coli* cells

E. coli TOP10 cells were streaked out on non-selective LB agar plates and incubated overnight at 37 °C. Chemically competent cells were generated by the Inoue method as described by Sambrook and Russell (2006). Electro-competent cells were prepared using a 40 mL pre-culture inoculated with a single colony and grown overnight at 37 °C and 130 rpm. The pre-culture was used to inoculate a 400 mL culture at an OD_{595} of 0.05. The cultures were grown at 37 °C and 130 rpm until an OD_{595} of 0.5 was reached and chilled in sterile centrifuge tubes on ice for 30 min with occasional mixing. The cells were centrifuged (4 °C; 10 min; 3,000 rpm) and washed with 200 mL sterile, ice-cold wash buffer (1 mM HEPES, 10 % (v/v) glycerol, pH 7.0). The centrifugation and washing step was repeated twice with sterile, ice-cold wash buffer (100 mL) and 10 % (v/v) glycerol solution (10 mL), respectively. After centrifugation, the cells were suspended 1 mL ice-cold, sterile 10 % glycerol solution and aliquoted in sterile 1.5 mL reaction tubes. The aliquots were frozen in liquid nitrogen and stored at -80 °C.

2.6.6 Transformation of *E. coli*

E. coli cells were typically transformed with 2 ng of intact plasmid DNA or ~10 ng of DNA from a ligation mix. For electroporation, the DNA was mixed with 50 - 60 µL of electro-competent cells in ice-cold 2 µm cuvettes and incubated on ice for 5 min. The transformation was performed with a Bio-Rad Gene Pulse™ according to the manufacturer's protocol. Immediately after the electroshock, 1 mL of SOC medium was added to the cells. Transformed cells were cultivated for 60 min at 37 °C with mild shaking. For chemical transformation, the DNA-cell mix was incubated on ice for 30 min. A heat shock was conducted at 42 °C for 30 s. Afterwards, the cells were cooled on ice for 2 min and 1 mL SOC medium was added. Cultivation was performed for 60 min at 37 °C with mild shaking. For both transformation methods alike, 50 to 200 µL of the bacterial suspension were plated on selective LB agar plates and incubated overnight at 37 °C.

2.6.7 DNA preparation and purification

2.6.7.1 DNA gel extraction and purification from reaction mixes

To extract DNA fragments from agarose gels, the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) was used according to the manufacturer's protocol. In addition, the kit was also used for direct purification of DNA from PCR and restriction digestion samples.

2.6.7.2 DNA plasmid mini and midi preparations

For DNA plasmid preparations, overnight cultures (16 h) of single *E. coli* clones carrying the respective plasmid were grown at 37 °C and 130 rpm in LB medium with appropriate antibiotics. Mini preparations were performed with the cells of 4 mL cultures with the GenElute™ Plasmid Miniprep Kit (Sigma) according to the supplier's manual with the following modifications. The elution was performed in 50 µL instead of 100 µL elution buffer. After the addition of the elution buffer, the columns were incubated for 1 min at room temperature prior to final the centrifugation step. Midi preparations were conducted with the cells of 200 mL overnight cultures using the PureYield™ Plasmid Midi Prep System Kit (Promega) according to the supplier's manual.

2.6.8 Generation and preparation of recombinant bacmids

Bacmids were created according to the Tn7 transposition based method with the EMBacY bacmid (Trowitzsch et al. 2010). Blue-white screening was used to identify clones carrying the recombinant EmBacY with the GOI. In this work, the vectors pFlpBtM-III-insect-Tmprss2-WT and pFlpBtM-III-insect-Tmprss2-D343N (§ 2.3) were used as donor vectors for the Tn7 transposition. The vectors were transformed into *E. coli* DH10(EmBacY) by electroporation (§ 2.6.6). The cells were selected on LB agar plates supplemented with tetracycline, gentamicin and kanamycin, bluo gal and IPTG. After growth, white colonies were picked and streaked out on fresh agar plates to avoid false positives. To prepare the recombinant bacmid, 20 mL of LB medium supplemented with kanamycin, gentamicin and tetracycline were inoculated from a single white colony. The culture was incubated at 37 °C and 130 rpm for up to 24 h. 1.5 mL of the culture were transferred to suitable reaction tubes and centrifuged for 2 min at 13,000 x g and 4 °C. The supernatant was completely removed by pipetting and the cells were resuspended in 300 µL of the resuspension solution P1. 300 µL of lysis solution P2 were added and mixed gently. The mix was incubated for 5 min at room temperature until it became translucent. 300 µL of the neutralization solution P3 were then added and the sample was incubated on ice for 10 min. The formed precipitate of protein and genomic DNA was pelleted by centrifugation for 10 min at 13,000 x g and 4 °C. The clear supernatant was transferred to a fresh reaction tube containing 800 µL of isopropanol to precipitate the bacmid DNA. The sample was mixed gently, incubated on ice for 10 min and centrifuged for 15 min at 13,000 x g and 4 °C. The supernatant was removed, the DNA pellet was washed with 500 µL of 70% ethanol and the centrifugation step was repeated. The ethanol was completely removed and the pellet was air dried at room temperature. Finally, the DNA was resuspended in 40 µL of the buffer NE from the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel) (5 mM TrisHCl, pH 8.5). The buffers P1-P3 are listed below in Table 2.33. The cells of the remaining 18.5 mL of the culture were used to generate 1 mL glycerol stocks in 10 % (v/v) glycerol. The vials were frozen in liquid nitrogen and stored at -80 °C.

Table 2.33. Buffers for the bacmid preparation.

Buffer	Composition
Solution P1 (Resuspension, Qiagen)	50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 µg/mL RNAse A
Solution P2 (Lysis, Qiagen)	0.2 M NaOH, 1 % (w/v) SDS
Solution P3 (Neutralization, Qiagen)	3 M Potassium Acetate (pH 5.5)

2.6.9 Preparation of electro-competent *P. pastoris*

Electro-competent *P. pastoris* cells were prepared and transformed according to the condensed protocol by Lin-Cereghino et al. (2005). The desired *P. pastoris* strain was streaked out on YPD agar plates without antibiotics (or appropriately selective media) and incubated at 30 °C for 48 h. A 50 mL pre-culture was inoculated and grown overnight in YPD at 30 °C and 130 rpm. A 400 mL YPD main culture was inoculated from the overnight culture at a final OD_{595} of 0.1 – 0.15 in a flask large enough to provide aeration. The culture was incubated at 30 °C and 130 rpm until an OD_{595} of 0.8 – 1.0 was reached. The cells were harvested for 5 min at 500 x g and the supernatant was removed. The cells were resuspended in 9 mL of ice-cold BEDS solution (10 mM bicine-NaOH, pH 8.3, 3 % (v/v) ethylene glycol, 5 % (v/v) dimethyl sulfoxide / DMSO) supplemented with 1 mL 1 M DTT. The suspension was incubated for 5 min at 30 °C and 100 rpm. The culture was centrifuged for 5 min at 500 x g and the cells were resuspended in 2 mL of BEDS solution without DTT. The competent cells were either used directly or frozen slowly in a Mr. Frosty™ Freezing Container with isopropanol (Thermo Fisher Scientific) and stored at -80 °C.

2.6.10 Stable transformation of *P. pastoris*

Prior to transformation into *P. pastoris*, the integrating vectors were linearized through unique digestion and subsequently purified (§ 2.6.3 and 2.6.7). The restriction enzymes used for linearization are listed in Table 2.34. Since the original CDS of Tmprss2 harbored a native *Sac*I site, *Dra*I was used to linearize the respective vectors. To generate single copy clones, 25 ng of linearized vector DNA were transformed. For multi copy clones, 100 ng of DNA were used. To perform the RMCE in P. pastoris, the vectors pBGP1Δ8K-FlpWT and pYEXUR-yeGFP were transformed in their circular states. The vectors were mixed prior to transformation with an excess of pBGP1Δ8K-FlpWT of 4:1 (80 ng : 20 ng). After the transformation, 1 mL 1 M D-Sorbitol solution was added to the cells. Following cultivation of the samples at 30 °C (mild shaking) for 1 h, the cells were plated on YPDS Zeocin agar plates or MD plates for appropriate primary selection. The plates were incubated at 30 °C for 2-3 d. For single clone isolation, grown colonies were diluted in 1 mL of sterile 1 M Sorbitol solution and streaked out again on fresh selection plates. To select for multi copy clones, the clones were streaked out on multiple plates with an increasing gradient of Zeocin (100 µg/mL to 1 mg/mL).

Table 2.34. Restriction enzymes for vector linearization to stably transform *P. pastoris*.

Homologous region	Vectors	Enzyme
***AOX1* Promoter**	All pPICZ(α) derived plasmids	*Sac*I / *Dra*I
***GAP* Promoter**	All pGAPZ(α)/pGAPZAΔ8 derived plasmids	*Avr*II
AOX1* Promoter and 3'*AOX1	pPIC9K, pYTAaox-RudolphRFP	*Bgl*II
URG	pYTAUR-RudolphRFP	*Sph*I

2.6.11 Genomic DNA extraction from *P. pastoris* and PCR analysis

The stable insertion of the vector into the genome of *P. pastoris* was analyzed by genomic PCR. The isolation of genomic DNA was performed adapted from the publication by Looke et al. (2011). *P. pastoris* clones were grown overnight from single clones at 30 °C and 130 rpm in appropriate medium. 100 - 200 µL of the overnight cultures (OD_{595} of ~0.4) were transferred to suitable reaction tubes and centrifuged at 12,000 x g for 1 min. The cells were resuspended in 100 µL of *P. pastoris* LiOAc lysis buffer (200 mM Lithium acetate, 1 % SDS) and incubated for 15 min at 90 °C. The lysate was briefly cooled on ice and 0.5 µL of RNase A (100 mg/mL stock, Qiagen) were added. The samples were incubated for 10 min at room temperature. 300 µL of ethanol p.a. or isopropanol were added and the sample was vortexed. Precipitation of the DNA could be extended to 30 min on ice or at -20 °C. The cell debris and the precipitated DNA were pelleted by centrifugation at 15,000 x g, 3 min. The pellet was washed with 1 mL 70 % ethanol (p.a.). After another centrifugation step (15,000 x g, 1min), the ethanol was carefully completely removed. The pellet was air dried and dissolved in 100 µL of elution buffer NE (NucleoSpin® Gel and PCR Clean-up Kit, Macherey Nagel). The dissolved DNA was separated from the cell debris by another centrifugation step at 15,000 x g, 30 s and transfer into a new reaction tube. The concentration was measured with the Nanodrop (§ 2.6.14). The PCR amplification of the genomic DNA was conducted as described in chapter 2.6.1. For each PCR, ~1 µL of the isolated genomic DNA were used. In order to exclude undesired background signal due to off-target binding of the PCR primers, the DNA of a respective control clone was used as a negative control template. The analysis of the *P. pastoris* RMCE master cell lines and the fluorochrome test cell lines was performed with two parallel PCRs. The first PCR (PCR1) was conducted to validate the integration of the vector into *P. pastoris*. The second PCR (PCR2) was conducted in parallel to distinguish single copy insertion clones from multi copy clones. A detailed overview of the PCR primers and the expected fragment sizes is described in Appendix II.

2.6.12 Functionality test of the selection trap Δ*his4* by cell growth

The selected clones were streaked out on YPD agar plates and grown for 48 h. One grown colony was dissolved in 1 mL of sterile 1 M Sorbitol solution. This "undiluted" mix was serially diluted to 1:100 and 1:1000 in 1 M Sorbitol. From all dilutions, 2.5 µL were dropped on MD and MDH agar plates. The plates were incubated for 48 h at 30 °C and the cell growth was documented.

2.6.13 Glycerol stocks of *P. pastoris*

Glycerol stocks were made from purified single clones, which were grown overnight at 30 °C and 130 rpm in 2 mL of suitable, selective medium. 800 µL of the culture were combined with 200 µL of sterile glycerol (p.a.) in 1 mL cryo vials and mixed by vortexing. The vials were frozen slowly in a Mr. Frosty™ Container with isopropanol (Thermo Fisher Scientific) and stored at -80 °C.

2.6.14 Quantification of DNA and protein concentrations

The concentration of DNA and proteins was assessed photometrically using a NanoDrop ND-2000c (Thermo Fisher Scientific) as described in the manufacturer manual. Nucleic acids were quantified at 260 nm. The $A_{260/280}$ ratio was used to track protein contaminations. A ratio of ~1.8 was accepted as pure DNA. A ratio of ~2.0 was accepted as pure RNA. Protein quantifications were carried out with the absorbance at 280 nm. To relate the absorbance to the actual protein concentration, it had to be divided by the molar extinction coefficient of the protein according to the Beer-Lambert law. The molar extinction coefficients of target proteins were calculated by Vector NTI. For recombinant Tmprss2, a factor of 0.45 mg/mL per 1 A_{280} was applied.

2.6.15 Transcriptional analysis by RT-PCR

2.6.15.1 Total RNA isolation from *P. pastoris* and cDNA synthesis

The clones were cultivated in 500 µL BMD medium in 2 mL 96-deepwell plates for 24 h at 28 °C and 1200 rpm in a Titramax 1000 incubator (Heidolph Instruments GmbH, Schwabach, Germany). Total RNA was isolated using NucleoSpin® RNA kit (Macherey Nagel), according to the supplier's manual. The cDNA was synthesized from the isolated total RNA using the peqGOLD M-MuLV H plus Reverse Transcriptase (VWR Life Science CC, Erlangen, Germany) according to the manufacturer's instructions. Protector RNase inhibitor, dNTP mix (PCR grade) and anchored Oligo$(dT)_{18}$ primer were taken from the Transcriptor First Strand cDNA Synthesis Kit (Roche Diagnostics, Mannheim, Germany). For each cDNA synthesis, a control was set up without reverse transcriptase (no reverse transcriptase, NRT) to exclude false positive signals during RT-PCR due to genomic DNA contamination.

2.6.15.2 Quantitative RT-PCR

The transcriptional analysis was conducted with four technical replicates of three independent biological clones of each tested cell line variant. The reverse transcription PCR (RT-PCR) was performed using the SsoFast™ EvaGreen® Supermix (Bio-Rad) according to the supplier's manual. The PCR was conducted for 40 cycles in a CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). Melting curves were analyzed to display only one peak. To exclude false positive

signals due to contamination of the PCR mixes, no template control (NTC) reactions were performed for each primer set. The RT-PCR was performed with three sets of primers. The names and expected PCR product sizes are depicted in Table 2.35. The sequences are listed in Appendix I.

Table 2.35. Used RT-PCR primers.

Name	Codes for	Function in Experiment	Primers	Fragment
RudolphRFP	RudolphRFP	Tested transcript of the GOI	rt_Rud-F rt_Rud-R	138 bp
PGK1	Phosphoglycerate kinase 1	Reference to normalize readouts of each sample	rt_PGK1-F rt_PGK1-R	137 bp
ACT1	Beta Actin	Reference, handled like GOI to evaluate comparability	rt_ACT1-F rt_ACT1-R	142 bpI

2.6.15.3 RT-PCR data analysis

Data analysis was performed using the comparative C_T (cycle threshold) method adapted from Schmittgen and Livak (2008). The readouts of each set of technical replicates were averaged. If no C_T readout was obtained during the experiment, the C_T was set by definition to 40. The first step comprised the "internal" calibration of each sample by correlating the readout for a tested GOI (*RudolphRFP* or *ACT1* as a control GOI) to the corresponding readout of reference gene (*PGK1*) by subtraction of the respective C_T values, resulting in the so-called ΔC_T value:

$$\Delta C_T = C_T(GOI) - C_T(Reference/PGK1) \quad \text{(Eq. 1)}$$

The ΔC_T values of each biological triplicate were averaged. The averaged ΔC_T values of all RMCE master cell line variants were put into relation by correlating them to the positive control (PC) as a selected reference sample by subtracting the respective ΔC_T values (resulting in the $\Delta\Delta C_T$ value). The relative fold change of the respective transcripts was calculated as $2^{-\Delta\Delta CT}$ for easier accessibility. For this, the PC was defined as 1 (= 100 %).

$$\Delta\Delta C_T = \Delta C_T(any\ tested\ cell\ line) - \Delta C_T(PC) \quad \text{(Eq. 2)}$$

$$Fold\ Change = 2^{-\Delta\Delta C_T} \quad \text{(Eq. 3)}$$

Student's t-tests were performed to ensure significance of the data. Gaussian Error Propagation was calculated as described by Papula (2008).

2.7 Cell culture techniques

2.7.1 Determination of cell number and viability

Cell numbers and viability (SF21, Hi5 and HEK293-6E) were assessed through the trypan blue dye exclusion method in a Neubauer hemocytometer. Trypan blue binds to proteins. As it is not able to permeate intact cell membranes, only dead cells are stained. To count the cells, 25 µL cell culture were diluted in 75 µL 0.5 % trypane blue solution. The stained and unstained cells of four big squares of the hemocytometer were counted. Cell density and viability were calculated as follows:

$$Cell\ Density\ \left(\frac{Cells}{mL}\right) = \frac{Cells\ \cdot\ 10^4\ \cdot\ dilution\ factor\ (4)}{number\ of\ counted\ big\ squares\ (4)} = Cells\ \cdot\ 10^4$$

$$Viability\ [\%]\ = \frac{Viable\ Cells\ \cdot\ 100}{Total\ Cells}$$

2.7.2 Cell size determination

The progress of baculoviral infection can be monitored through an increase in the diameter of infected insect cells. This was done with the CASY Cell Counter (Innovatis), which delivers simultaneous readouts for the number, the viability and the size of measured cells following a cell type specific calibration. The device was used according to the manufacturer's guidelines after the samples were diluted in an isotonic, particle pure buffer to reach the measurement range of the device. The CASY counter works by electric current exclusion. The measured particles (cells) are diluted in an electrolyte and flow through a capillary passing an electrical field, which induces a resistance that is proportional to the size of the measured cells. This also allows measuring the viability of a sample, as the intact membrane of living cells exhibits a significantly higher resistance compared to dead cells.

2.7.3 Analytic flow cytometry with the Guava EasyCyte Mini

In this work a Guava EasyCyte Mini System (Merck) was used to monitor the efficiencies of the transfections of HEK293-6E cells as well as the baculoviral infections of SF21 and Hi5 cells. Moreover, the expression analysis of the *P. pastoris* RMCE master cell line variants and exchanged producer cell lines was conducted by flow cytometry to quantify the expression of RudolphRFP and y-eGFP. The Guava EasyCyte Mini is equipped with a 488 nm argon laser for excitation. It offers three band-pass filters (525/30, 583/30 and 680/30) to detect the emission of green, yellow and red signals. In this work, the 525/30 filter was used to detect the expression of GFPq, y-eGFP and YFP. RudolphRFP was detected with the 583/30 band-pass filter. For the determination of fluorescence the cell suspensions were diluted 1:10 in PBS. For the expression analyses in *P. pastoris*, the mean red or green fluorescence was defined as the measurement value.

2.7.4 Transient transfection of HEK293-6E with PEI

Serum-free cultivated HEK293-6E cells were transfected using linear polyethylenimine (PEI) with an average molecular weight of 25 kDa (Polysciences). The cells were transfected in their exponential growth phase at a cell density of 1.5 – 2.0 x 10^6 cells/mL. The transfected DNA comprised a mix of 95 % expression vector (pFlp-BtM-III-Tmprss2-D343N or pFlpBtM-III-Tmprss2-WT) and 5 % of pTTO/GFPq, which was used to monitor the transfection efficiency. Per 30 mL culture (set up in fresh medium), the transfection mix was prepared with 30 µg of the DNA mix and 75 µg of PEI. The DNA and PEI were each separately added to 1.5 mL of F17 medium in sterile 15 mL polystyrene tubes and vortexed briefly. Afterwards, the DNA solution added to the PEI solution and the mix was vortexed briefly. After incubation at room temperature for 15 min, the now complexed transfection mix was added to cell culture.

2.7.5 Transfection of SF21 cells with bacmids and virus amplification

The generation of baculoviruses to express Tmprss2-WT and Tmprss2-D343N was performed by transfecting adherent SF21 cells with the generated recombinant EMBacY bacmids (§ 2.6.8) using SuperFect (Qiagen). For the preparation of the first virus generation 0.4 x 10^6 SF21 cells from a culture supplemented with 5 % FCS were seeded on 6-well plates. For each transfection, 5 µL of the isolated recombinant bacmid and 10 µL of SuperFect were diluted in 150 µL serum free Ex-Cell 420 medium for pre-complexing. The mix was incubated for 10 min at room temperature and supplemented with 850 µL of Ex-Cell 420 with 5 % FCS. This final mixture was used to substitute the culture media covering the adherent cells in the 6-well plates. After 2 h of incubation at 27 °C, further 1.5 mL of Ex-Cell 420 with 5% FCS were added to each well. The initial virus V0 was harvested with the supernatant after 3 - 5 d of incubation at 27 °C, depending on the YFP expression from EmBacY. The baculovirus-containing supernatant was used to perform one round of virus amplification (VA1). For this, serum-free suspension SF21 cells were seeded in a density of 0.5 × 10^6 cells/mL in 2 x 50 mL cultures and incubated at 27 °C and 100 rpm. The cultures were then infected through the addition of 2 % (v/v) of the harvested virus supernatant. The cell number, viability and cell diameter were determined every 24 h by manual counting and with the CASY Counter as described in chapters 2.7.1 and 2.7.2. Furthermore, the expression of YFP was monitored using the Guava Flow EasyCyte System (§ 2.7.3). The virus-containing supernatant was harvested 48 h after the cells had stopped proliferating. The supernatant was separated from the cells by centrifugation at 5,000 x g for 15 min and filtered with 0.45 µm diameter. The cleared VA1 was directly used to infect SF21 and Hi5 expression cultures.

2.8 Protein production and purification

2.8.1 Stable protein expression in *P. pastoris*

The cultivation of the generated and purified *P. pastoris* producer clones were conducted at 28 °C to 30 °C. Shake flask cultures were performed in baffled flasks at 130 rpm. High-throughput expression screenings using 2 mL 96-deepwell plates were conducted at 1000 to 1200 rpm in a Heidolph Titramax 1000 incubator (with at most 1.5 mL per well). For comparative expression experiments, the OD_{595} was tested to ensure the comparability of the results. In all experiments except for high throughput expression in 96-deepwell plates, 1 mL samples were taken every 24 h. The cells were spun down at 14,000 x g for 3 min. For secretory proteins (Tmprss2, HSA) the supernatant and the cells were stored separately at -20 °C after freezing in liquid nitrogen for later analyses. For intracellular proteins (fluorochromes) the supernatant was discarded and only the cells were stored.

The *GAP* promoter based expression in GS115 was performed in YPD, BMD or BMDY medium. The expression was conducted for 48 h, unless stated differently. During comparative analyses, technical replicates of a single biological clone were inoculated from one common pre-culture. The expression was started at an inoculated OD_{595} of 0.1. For the *AOX1* promoter based expression in GS115 (mut^+), the cells were grown in BMGY for 24 h and resuspended in an equal volume of BMMY after centrifugation at 3,000 x g for 5 min to induce the expression. The expression was conducted for 48 hpi (hours post induction). The induction was maintained by the addition of 0.5 % (v/v) methanol every 24 hpi.

For the *AOX1* based expression in the mut^S strain KM71H pre-cultures were grown in BMGY medium for 24 to 48 h or until an OD_{595} of ~10 was reached. The cells were harvested at 3,000 x g for 5 min. The supernatant was discarded. In shake flasks, cells were resuspended at 1/5 of the pre-culture volume in BMMY medium. In 96-deepwell plates, 1/3 of the preculture volume was used instead for easier handling. The expression was conducted for 48 hpi, unless noted otherwise. The induction was maintained by the addition of 0.5 % (v/v) methanol every 24 hpi. In later experiments, the antifoam Struktol® J 673 A was added at 1 % (v/v) to the culture media. To harvest the supernatant after the expression phase, the cultures were centrifuged at 6,000 x g for 10 min. The supernatant was subsequently fully cleared by to two more rounds of centrifugation at 24,000 x g for 30 min and passage through a 0.2 µm filter.

The *AOX1* promoter based expression in KM71H was scaled up to 2 L or 5 L Labfors bioreactors (Infors HT, Bottmingen-Basel, Switzerland). The expression was conducted according to the protocol of the Helmholtz Protein Sample Production Facility (PSPF) at the HZI (Braunschweig, Germany) with the media described in Table 2.11 – Table 2.15 (§ 2.1.3). The cultivation was

conducted at 28 °C and ~pH 6.0. The pH was maintained with 12.5 % (v/v) ammonia solution. Stirring (6-bladed impeller) was started at 400 rpm and increased if necessary to ensure aerobic culture conditions ($pO_2 > 0$). The process parameters were monitored during the run. To start a reactor run, a first pre-culture of 50 mL was grown in shake flasks with YPD medium, supplemented with Zeocin (100 µg/mL). From this culture, a 400 mL YPD pre-culture was inoculated. Once the culture reached an OD_{595} of 15 to 20, it was used to inoculate the installed and sterilized bioreactor at a final OD_{595} of 1. The cells were grown for ~24 h, until the glycerol of the bioreactor growth medium was used up, which was monitored through an increase of the pO_2. The cultivation was then continued at carbon-source limiting conditions through the slow addition of the glycerol feed solution to the reactor. As soon as an OD_{595} of 180 to 250 was reached, the glycerol feed was stopped and the exhaust of glycerol in the medium was again tracked by the pO_2. At this point, methanol feeding was started. The cells were slowly adapted to methanol by its step-wise addition of 0.1 % – 0.5 %. A methanol sensor and an automatic pump (Raven Biotech Inc., Vancouver, Canada) were used to maintain 0.5 % (v/v) methanol in the bioreactor. The expression was conducted for 48 hpi. The culture supernatant was harvested by centrifugation at 6,000 x g for 10 min, separated from the cells and fully cleared by to two more rounds of centrifugation at 24,000 x g for 30 min followed by passaging through a 0.2 µm filter.

2.8.2 Plasmid based transient expression in HEK293-6E

Transfections were carried out as described in chapter 2.7.4. The transformed cells were cultivated under standard conditions (37 °C, 100 rpm, 5 % CO_2 in a humidified atmosphere). The cell count, the viability the green fluorescence of GFPq were analyzed every 24 h. Samples of the culture supernatant as well as aliquots of 1 x 10^6 cells were taken for analysis and stored at -20 °C until further use. The transfected cultures were expanded 48 h post transfection (hpt) with fresh medium to a total volume of 50 mL and supplemented with tryptone (TN1) at a final concentration of 0.5 %. After 72 hpt the cultures were additionally fed with 4.5 g/L glucose. The expression was terminated after 96 hpt and the total culture was harvested for further analysis. The cells were separated from the supernatant by centrifugation at 500 x g for 5 min.

2.8.3 Protein expression in SF21 and Hi5 insect cells using BEVS

The virus amplification was carried out as described in chapter 2.7.5. For baculoviral protein production insect cells were seeded in 120 mL suspension cultures of each SF21 and Hi5 cells with an initial cell density of 1x 10^6 cells/mL. The cells were infected with 10 % (v/v) of VA1. The infection kinetics were tracked every 24 h through cell count, cell diameter and the percentage of the YFP fluorescent cells. Samples of the culture supernatant as well as aliquots of 1 x 10^6 cells were taken for analysis and stored at -20 °C until further use.

The cultivation was stopped at 72 h post infection for SF21 and at 96 h post infection for Hi5. The whole culture was harvested and the cells were separated from the supernatant by centrifugation at 500 x g for 5 min.

2.8.4 Cell lysis of *P. pastoris*

The cell lysis of *P. pastoris* was carried out with acid washed glass beads (size 0.5 mm) according to the Pichia Expression Kit manual (Thermo Fisher Scientific). The cell pellets of harvested 1 mL samples were thawed quickly and placed on ice. 100 µL of ice cold Breaking Buffer (BB, Table 2.36) were used to resuspend each cell pellet. The volume was estimated and acid-washed glass beads were added roughly at the same volume. The cell-bead mixtures were subjected to eight cycles of rigorous vortexing and incubation on ice for respectively 30 s. The samples were centrifuged for 10 min at 16,000 x g and 4°C. The clear soluble lysate fraction was transferred to fresh 1.5 mL reaction tubes. To obtain the insoluble fraction, an equal amount of fresh lysis buffer was added to the remaining mix of glass beads and cell debris. To prepare SDS-PAGE samples, 50 µL of lysate were mixed with an appropriate volume of SDS-loading buffer (§ 2.9.1,Table 2.43) and boiled for 5 min at 95 °C.

Table 2.36. *P. pastoris* Breaking Buffer (BB).

Component	Final concentration
Sodium phosphate buffer, pH 7.4 (1 M stock)	50 mM
EDTA	1 mM
Glycerol p.a.	5 % (v/v)
in MQ-H_2O	
To add before use	
DNAse (1 mg/mL)	2 - 5 µg per 50 mL of BB
Protease Inhibitor (Roche cOmplete™, EDTA-free)	1 tablet per 50 mL of BB

2.8.5 Cell lysis of HEK293-6E, SF21 and Hi5 cells

250 µL of lysis buffer (Table 2.37) were added to the cell pellet of harvested 1 mL samples. The cells were resuspended by vortexing. The mixture was kept on ice for 15 min and centrifuged at 16,000 x g for 15 min to separate the soluble and insoluble fractions. For the lysis of the total cultures, 25 mL of lysis buffer were added to the frozen cells. The mixture was vortexed rigorously and the tube was then incubated on ice for approximately 30 min. The resuspended cells were disrupted by sonication in a Sonopuls sonifier (Bandelin) using 80 W in intervals of 1 s for 2 min. The samples were centrifuged at 30,000 x g for 30 min in order to separate soluble and insoluble fractions. The samples could be stored at -20 °C until the analysis was conducted.

Table 2.37. Mammalian and insect lysis buffer for protein extraction.

Component	Final concentration
Sodium phosphate buffer, pH 8.0	50 mM
Imidazole	5 mM
NaCl	300 mM
IGEPAL CA-630	0.5 % (v/v)
Beta-mercaptoethanol	3 mM
in MQ-H_2O	
To add before use	
DNAse (1 mg/mL)	5 - 10 µg per 50 mL of lysis buffer
Protease Inhibitor (Roche cOmplete™, EDTA-free)	1 tablet per 50 mL of lysis buffer

2.8.6 Trichloroacetic acid (TCA) precipitation of proteins

TCA precipitation was used as a simple method to concentrate small samples for tests of protein expression by SDS-PAGE and Western blot. 1 mL of culture supernatant or total cell lysate was mixed with 100 µL of 100 % TCA. The mixture was incubated on ice for 30 min and centrifuged at 15,000 x g and 4 °C for 15 min to pellet the precipitated protein. The protein pellet was washed in 1 mL of ice-cold 70 % ethanol and centrifuged again for 5 min at 15,000 x g and 4 °C. Remaining ethanol was removed and the protein pellet was dissolved in 50 µL NE buffer (5 mM Tris-HCl, pH 8.5) from the NucleoSpin® Gel and PCR Clean-up Kit (Macherey-Nagel).

2.8.7 Dialysis and diafiltration

Dialysis was performed to exchange the buffer of small volumes (< 50 mL per sample) using dry Spectra/Por® dialysis tubes (Spectrum Laboratories Inc., Rancho Dominguez, California, USA) with a MWCO of 10 kDa or less. Prior to use, the tubes were soaked in the desired exchange buffer. Dialysis was conducted in two repetitions at 4 °C with slow stirring of the buffer with at least volumes of exchange buffer compared to the volume of the original sample. Diafiltration of 50 mL to 500 mL culture supernatant were conducted using the VIVAflow 200 system (Sartorius AG) with a 10 kDa MWCO membrane according to the supplier's manual. Prior to use, the 10 % storage ethanol solution was completely removed from the cartridge. Thereafter, the cartridge was washed with least 500 mL of MQ-H_2O to remove residual ethanol. The water was completely drained from the system. The supernatant was initially concentrated in the device. 200 mL of the supernatant were placed inside the concentrate reservoir. The remaining supernatant was connected to the reservoir from a separate bottle (feed reservoir). The sample was concentrated to ca. 50 mL to 100 mL by circulation through the system. The concentrated sample was dialyzed by connecting the desired exchange buffer to the feed reservoir and continued recirculation of through the system with least 10 volumes of buffer. After recovery of the concentrated and diafiltrated sample, the system was cleaned according to the supplier's guidelines and stored in 10 % ethanol. Large-scale diafiltration (> 500 mL) was performed with a ProFlux® tangential flow system (Merck) equipped with two Pellicon 2 cassettes (Merck) with 10kDa MWCO according to the manufacturer's

instructions. Samples were concentrated to ~500 mL and diafiltered with at least 10 volumes of buffer. The concentration and diafiltration of cleared and filtered *P. pastoris* culture supernatants in the ProFlux unit was done by Nadine Konisch (RPEX, HZI).

2.8.8 Concentration of protein samples with Vivaspin concentrators

Vivaspin concentrators (Sartorius AG, Göttingen, Germany) with polyether sulfone (PES) membranes were used to concentrate smaller volumes. For this, protein solutions were centrifuged at < 5,000 x g (Vivaspin 20) or < 4,000 x g (Vivaspin 2 and Vivaspin 6) until the desired volume was reached. Buffer exchange could be performed by refilling the columns with the exchange buffer after concentration. This was done with at least 10 volumes (related to the original sample).

2.8.9 Small-scale protein capture with magnetic beads

Small-scale His tag batch purifications were performed with MagneHis™ Ni-Particles (Promega) according to the supplier's manual with 30 µL of uniformly suspended magnetic beads per 1 mL of dialysed or diafiltered *P. pastoris* culture supernatant. All reactions were performed in 100 mM HEPES buffer pH 7.5, 500 mM NaCl. For the washing buffer, imidazole was added at a final concentration of 10 mM. The elution buffer was supplemented with 500 mM imidazole. For binding, the mix was gently inverted ten times and incubated for 5 min at room temperature while keeping the material suspended. Three washing steps were performed, before the protein was eluted into 100 µL elution buffer through 5 min of batch incubation.

2.8.10 Small-scale batch binding analysis with resin material

The analysis was performed in 50 mL reaction tubes with 50 µL of equilibrated beads of either Ni-NTA His-Bind® Superflow™ Resin (Novagen) against a His tag or of ANTI-FLAG® M2 Affinity Gel (Sigma) against a Flag tag. Both materials were used following the supplier's protocols with 50 mM sodium phosphate buffer pH 7.4, 300 mM NaCl serving as the binding buffer. From a harvested and cleared, but not diafiltered, culture supernatant, four 20 mL samples were taken to test the two resin materials respectively under two conditions. In the first step, all samples were equally supplemented with a final concentration of 50 mM sodium phosphate buffer pH 7.4, 300 mM NaCl to adjust the pH for binding. The binding to Ni-NTA material was tested with or without the addition of 1 mM $NiSO_4$. For the ANTI-FLAG material, the elution at low pH (pH 3.5) and using Flag peptide (SBAU, HZI) were compared. An overview is shown in Table 2.38. After sample preparation and initial probing (SDS-PAGE samples), equilibrated resin materials were added. Binding was performed overnight at 4 °C with gentle shaking. The material was separated from the supernatant by centrifugation at 500 x g for 5 min. The supernatant was now treated as the "non-bound" fraction (= "flowthrough"). The beads were transferred to fresh

reaction tubes for easier handling, washed 3x by resuspension in 1 mL of the appropriate wash buffer followed by centrifugation at 500 x g for 1 min. Elution was performed under the conditions denoted in Table 2.38. For the low pH elution, the pH of the eluate was immediately neutralized by the addition of a final concentration of 25 mM Tris-HCl buffer, pH 8.0. SDS-PAGE samples were prepared from all liquid fractions (§ 2.9.1) and from the resin material. For this, the material was suspended in 2x concentrated SDS loading buffer after the elution and boiled at 95 °C for 5 min.

Table 2.38. Overview of the tested conditions for direct capture of Tmprss2 by Ni-NTA or ANTI-FLAG.

Sample	Modified binding conditions	Elution conditions
Ni-NTA 1	-	Competitive (500 mM imidazole)
Ni-NTA 2	Presence of 1 mM $NiSO_4$	Competitive (500 mM imidazole)
α-Flag 1	-	Acid (0.1 M glycine HCl, pH 3.5)
α-Flag 2	-	Competitive (100 µg/mL Flag peptide)

2.8.11 Large-scale protein capture by affinity chromatography

The secretory Tmprss2 constructs expressed in this work were equipped with a C-terminal 8xHis tag that was mainly used for primary affinity capture on nickel or cobalt based resin material. The capture was performed with self-packed Mobicol columns (MoBiTec) or an ÄKTA-FLPC unit (GE Healthcare). The recipes of the buffers used in this work are listed in Table 2.39. The elution was performed competitively against imidazole. For this, two buffers (A and B) were prepared with different imidazole concentrations. Buffer A was prepared without imidazole ("0 mM") and directly used as the binding/equilibration buffer. Buffer B was prepared with 1 M of imidazole. His wash buffer (10 mM of imidazole) and His elution buffer (150 - 500 mM imidazole, depending on the resin material) were prepared by mixing buffers A and B. Unless noted otherwise, all supernatants were diafiltered against binding buffer and concentrated to 500 mL prior to loading onto the column (§ 2.8.7). Buffer B could be added to the diafiltered supernatant at a final concentration of 10 mM imidazole to enhance the specificity of the Ni-NTA binding.

For manual His capture with self-packed columns, Ni-NTA His-Bind® Superflow™ Resin (Novagen) was used. The experiment was carried out according to the manufacturer's instructions. A 20 mL syringe was screwed on top of the stacked column as a reservoir. Prior to use, the resin was washed with 10 column volumes (CV) each of 20 % ethanol and MQ-H_2O. A third washing step was conducted with buffer A. The supernatant was loaded onto the resin at appropriate speed for the resin material and the size of the column using a peristaltic pump. After loading, the resin was washed with wash buffer, until no signal was obtained anymore from an attached and calibrated UV monitor (280 nm). The target protein was eluted with elution buffer in a series of 500 µL to 1 mL fractions. The elution was also monitored by UV (280 nm). The affinity capture with an ÄKTA-FPLC (GE Healthcare) was performed with prepacked HisTrap FF or HiTrap TALON® columns of 1 mL or 5 mL size (GE Healthcare). The ÄKTA unit was operated according

to the manufacturer's instructions. Buffers A and B were degassed by ultra-sonication prior to use. The eluent fraction was collected in a series of 1 mL samples using an automatic sampler.

Capture by ion exchange was conducted with Q-Sepharose Fast Flow resin (GE Healthcare) according to the manufacturer's instructions. The buffers are listed below in Table 2.39. The capture of Tmprss2 was performed at pH 9.0 with a gradient of 0 - 500 mM NaCl in an ÄKTApilot (GE Healthcare) that was operated as instructed by the manufacturer. A XK 50/30 column was packed with 25 mL of resin material and equilibrated prior to use. The cleared *P. pastoris* culture supernatant was diluted 1:10 in Q-Sepharose FF Buffer A before being added to the column. The eluent fraction was collected in a series of samples in accordance to the monitored UV 280 nm absorbance. The ion gradient was monitored through the conductivity.

For every affinity capture, the different fractions of the prufication process were analyzed by SDS-PAGE and Western blot (§ 2.9.1, 2.9.3).

Table 2.39. Buffer recipes for protein Ni-NTA and Q-Sepharose capture.

Buffer	Composition
His Buffer A (variant 1)	50 mM sodium phosphate, pH 7.4, 150 – 300 mM NaCl
His Buffer B (variant 1)	50 mM sodium phosphate, pH 7.4, 150 – 300 mM NaCl, 1 M imidazole
His Buffer A (variant 2)	50 mM Tris-HCl, pH 7.4 – 8.0, 150 – 300 mM NaCl
His Buffer B (variant 2)	50 mM Tris-HCl, pH 7.4 – 8.0, 150 – 300 mM NaCl, 1 M imidazole
Q-Sepharose FF Buffer A	25 mM Tris-Base, pH 9.0 (0 mM NaCl)
Q-Sepharose FF Buffer B	25 mM Tris-Base, pH 9.0, 1 M NaCl

2.8.12 Deglycosylation by Endo Hf

Endo Hf (NEB, Ipswich, Massachusetts, USA) (1×10^6 U/mL) was used according to the manufacturer's guidelines. Both native and denaturing conditions were respectively tested with ~1 µg of Tmprss2-D343N. For denaturing conditions, the protein was mixed with 1 µL 10x Glycoprotein Denaturing buffer and MQ-H_2O to a final volume of 10 µL. The sample was incubated for 10 min at 100 °C. Thereafter, 2 µL of 10x GlycoBuffer 3, 3 µL H_2O and 5 µL of Endo Hf were added to make a final reaction volume of 20 µL. For native conditions, the first two steps were skipped and the 20 µL reaction mix was directly set up with the protein and consequently a higher volume of MQ-H_2O. The deglycosylation reaction was conducted for 1 h at 37 °C. The recipes of the reaction buffers are noted below.

Table 2.40. 10x Deglycosylation buffers (Endo Hf).

Buffer	Composition
10x Glycoprotein Denaturing Buffer	5 % (v/v) SDS; 40 mM DTT
10x GlycoBuffer 3	50 mM sodium acetate; pH 6.0

2.9 Protein analytical methods

2.9.1 SDS-PAGE

SDS-PAGE (sodium dodecyl sulphate polyacrylamide gel electrophoresis) was used to analyze protein quality and yields under denaturing conditions (Laemmli 1970). Self-made gels were used consisting of a lower resolving gel (10 % or 12 % acrylamide) and an upper stacking gel (5 % acrylamide) (Table 2.41; Table 2.42). Usually 15 µL to 30 µL of protein sample were mixed with an appropriate amount of 4 x SDS loading buffer (Table 2.43) and heated for 5 min at 95 °C for denaturation. To enhance resolution, the samples were collected in the stacking gel by an initial run at 120 V for 5 min. Thereafter, the gels were run at 160 V for ~1 h, until the bromphenol blue front reached the end of the gel. The SDS gel was washed 3x 5 min in MQ-H_2O to remove excess SDS and subsequently stained with either InstantBlue (Expedeon, San Diego, USA) or Coomassie staining solution (Table 2.41). Alternatively to self-cast gels, Criterion™ precast gels (Bio-Rad) were used according to the supplier's protocols. The electrophoresis was then performed at 150 V. The used molecular weight standards are listed in Table 2.2.

Table 2.41. Buffers and reagents for SDS-PAGE.

Buffer	Composition
4× Upper buffer	0.5 M Tris-HCl pH 6.8, 0.4 % SDS (v/v)
4× Lower buffer	1.08 M Tris-base, 0.42 M Tris-HCl, pH 8.8
SDS Running Buffer	3 g/L Tris-base, 14.4 g/L Glycin, 1 g/L SDS, pH 8.3
Coomassie Staining	50 % Ethanol, 10 % acetic acid, 0.25 % Coomassie R-250
Coomassie Destaining	50 % Ethanol, 10 % acetic acid

Table 2.42. Compositions of SDS-PAGE gels (for eight gels).

Component	12 % Resolving Gel	10 % Resolving Gel	5 % Stacking Gel
Acrylamide/bisacrylamide 30 % (v/v)	12 mL	10 mL	1.5 mL
4× Upper buffer	-	-	2.5 mL
4× Lower buffer	7.5 mL	7.5 mL	-
10 % SDS solution (in MQ-H_2O)	0.3 mL	0.3 mL	-
MQ-H_2O	10.1 mL	12.4 mL	5.9 mL
TEMED	40 µL	40 µL	30 µL
40 % APS	60 µL	60 µL	30 µL

Table 2.43. 4x SDS loading buffer.

Component	Final concentration
SDS	7 % (v/v)
Glycerol p.a.	0.22 % (v/v)
Tris-HCl pH 6.8	9.6 mM
Beta-mercaptoethanol	2.4 mM
Bromphenol blue	0.22 mg/mL
in MQ-H_2O	

2.9.2 Native PAGE

In a native PAGE, protein samples are separated without denaturation of secondary structures. Hence, the migration is influenced by the net charge on the protein surface and the hydrodynamic radius in addition to its mass. To ensure protein migration to the cathode in the absence of SDS, a slightly negative charge is applied through Coomassie G-250, which does not denature the proteins (Schägger and von Jagow 1991). The NativePAGE™ Novex Bis-Tris Gel System (Thermo Fisher Scientific) was used (4-16% gels) in this work according to the manufacturer's instructions. The HMW Native (66 - 669 kDa) and LMW Native (14.4 - 97 kDa) standards by GE Healthcare were used as molecular weight markers.

2.9.3 Western blotting

The sensitive detection of specific proteins from culture supernatant was performed by Western blots and immunostaining. Either a Trans-Blot-SD device or a Trans-Blot Turbo Transfer System (both by Bio-Rad) were used to mediate the transfer of protein samples from a freshly run SDS polyacrylamide gel to a polyvinylidene difluoride (PVDF) membrane (Immobilon-P, Merck) in a semi dry blotting procedure. The SDS gel and two gel-sized pieces of Western blot filter paper were equilibrated for 10 min in Western blot transfer buffer (25 mM Tris-base, pH 8.0, 192 mM Glycine, 15 % (v/v) methanol). The PVDF membrane was briefly activated with methanol and subsequently equilibrated in transfer buffer for 5 min. The gel was directly placed on the membrane. Both were then allocated between two layers of the soaked filter paper and placed on the anode of the blotting apparatus. The blot was run at 14 V for 30 min (Trans-Blot-SD) or for 7 min with mixed Mw program (Trans-Blot Turbo). Unspecific binding sites for antibodies on the PVDF membrane were saturated/blocked by incubating the membrane at room temperature for 30 - 60 min in 10 mL of TBS-T (20 mM Tris-base, pH 8.0, 150 mM NaCl, 0.05 % (v/v) Tween) supplemented with 2.5% (w/v) of skim milk powder (blocking solution). The saturated membrane was washed 3x for 5 min in TBS-T and incubated at 4 °C overnight (mildly shaking) with 10 mL of TBS-T supplemented with an appropriate amount of primary antibody. If the primary antibody was compatible to blocking solution, the washing steps following the saturation of the membrane were skipped and the antibody was directly applied in 10 mL of fresh blocking solution. After antibody-binding, the membrane was washed 3x for 5 min in TBS-T and incubated with an appropriate secondary antibody, conjugated either to AP (alkaline phosphatase) or HRP (horseradish peroxidase) in TBS-T. The binding was performed at room temperature for 2 h. The membrane was washed 3x 5 min in TBS-T to remove the unbound secondary antibodies. For AP staining, the membrane was equilibrated for 5 min shaking in AP buffer (100 mM Tris base pH 9.5, 100 mM NaCl, 5 mM $MgCl_2$). Staining was performed by the BCIP/NBT Color Development Substrate (Promega) with 33 µL BCIP (50 mg/mL) and 66 µL NBT (50 mg/mL) in 10 mL AP buffer. After a

maximum of 30 min the reaction was terminated in MQ-H_2O. HRP staining was performed using the Lumi-Light Western Blotting Substrate (Sigma) according to the supplier's manual. The protein was detected in a Fujifilm LAS-3000 Imager.

The antibodies used in this work are listed in Table 2.44. If not stated otherwise, the detection of Tmprss2 constructs was performed in this work with a mouse-produced ANTI-FLAG M2 primary IgG monoclonal antibody (Sigma) and an AP-conjugated goat anti-mouse secondary IgG polyclonal antibody (Promega).

Table 2.44. Antibodies used for Western blots.

Antibody	Type	Working Concentration	Supplier
Mouse α-Flag	monoclonal primary	1:4000 in blocking solution or TBS-T	Sigma (F3165)
Mouse IgG1 α-His-Tag	monoclonal primary	1:2000 in TBS-T	Merck (#70796)
Goat α-mouse IgG (H+L)-AP	polyclonal secondary	1:7500 in TBS-T	Promega (S3721)
Goat α-mouse IgG(H+L)/IgM-HRP	polyclonal secondary	1:3000 in TBS-T	Dianova (115-035-062)
Rabbit α-PR8HA (H1)	polyclonal primary	1:4000 in blocking solution or TBS-T	Sino Biological (11684)
Goat Anti-Rabbit IgG (H+L)-AP	polyclonal secondary	1:5000 in TBS-T	Bio-Rad (#1706518)

2.9.4 Slot blotting

Slot blots allow loading liquid samples containing the target protein directly onto a nitrocellulose or PVDF membrane for immunostaining. In contrast to Western blotting, the slot blot allows to quickly use the sensitivity of immunostaining without the need to perform SDS-PAGE first. Since the transfer is usually realized by pulling the sample liquid through the membrane with applied vacuum, it is possible to load greater amounts of the sample into one concentrated spot on the membrane compared to SDS-PAGE and Western blot. However, the samples are not separated by size on a slot blot. Also, staining due to cross-reactivity of the applied antibody (background/noise) cannot be separated from the true signal. Hence, slot blotting only comprises a preliminary expression screening to compare different clones and select the most promising ones for further analyses. In this work, slot blots were performed with a Minifold® II Slot-Blot Manifold System (Cole-Parmer, Vernon Hills, Illinois, USA) for high throughput expression screenings of Tmprss2. The apparatus was prepared according to the manufacturer's guidelines with a PVDF membrane (Immobilon-P, Merck). The preparation of the membrane followed the same method (including buffers and solutions) described in chapter 2.9.3. 150 to 300 µL of *P. pastoris* culture supernatant were loaded per slot. Vacuum was only applied as briefly as possible, until the samples had been completely pulled through. Immunostaining was performed as described in chapter 2.9.3.

2.9.5 Mass spectrometry

MALDI-TOF (Matrix Assisted Laser Desorption Ionization – Time-Of-Flight) mass spectrometry (MS) was used to identify protein fragments of SDS-PAGE samples. To prepare the samples for tryptic digestion, the corresponding bands were cut out of stained SDS gels and transferred into 1 mL of MQ-H_2O. After a tryptic digest, proteins were co-crystallized with organic acids. Desorption of protein fragments was triggered by a UV laser beam. The TOF measurements exposed mass-to-charge ratios of the ions. This data was compared with the MASCOT database to identify the analyzed proteins. All MS analyses were conducted by MS platform of the research group Cellular Proteomics (CPRO) at the HZI (Braunschweig, Germany).

For intact mass analyses by ESI (Electrospray Ionization), the buffer of purified protein samples was changed to MQ-H_2O to remove any residual salt by 3x complete buffer exchange on Viva Spin 500 µL columns (Sartorius). The samples were immediately handed in for MS analysis.

2.9.6 Enzymatic *in vitro* activity test of recombinant Tmprss2

The enzymatic activity of Tmprss2-WT and Tmprss2-D343N was analyzed through a preliminary activity assay *in vitro*. Recombinant hemagglutinin H1 produced in Hi5 cells (produced by Margitta Schürig, research group Recombinant Protein Expression, HZI) was used as a substrate for the reaction. 10 ng of either Tmprss2-D343N or Tmprss2-WT were mixed with 10 µg of H1 and incubated in 50 mM Tris-HCl buffer, pH 7.4 with 150 mM NaCl at 37 °C. Control reactions without Tmprss2 were performed in parallel. Samples were taken after 3 h and 16 h and boiled in SDS sample buffer and SDS-PAGE was performed (§ 2.9.1). The samples were analyzed by Western blot (§ 2.9.3). H1 was detected with a rabbit α-PR8HA (H1) primary antibody and a goat α-rabbit AP-coupled secondary antibody (§ Table 2.44).

3 Results

3.1 Establishment of a RMCE system in *P. pastoris*

The generation of *P. pastoris* producer cell lines by homologous recombination is usually a time-intensive task, because depending on the target protein, extensive expression screens are required to identify a high producer ("jackpot") clone. The recombination mediated cassette exchange (RMCE) allows for the introduction of any gene of interest (GOI) into a specifically "tagged" genomic locus of a RMCE master cell line (MCL), which harbors an exchangeable RMCE cassette. This should result in clones with uniform product yields. Moreover, these product yields should scale with the expression of a model GOI / marker protein present inside the RMCE tagging cassette before the cassette exchange. Hence, the RMCE presents a promising approach for the fast generation of *P. pastoris* producer cell lines with predictable expression results.

For this reason, it had to be evaluated, whether the RMCE system could be established in *P. pastoris* through an initial model system. Furthermore, it had to be tested, whether the cassette exchange would indeed result in producer clones with comparable expression levels of the GOI.

3.1.1 Experimental design of the RMCE system in *P. pastoris*

The initial aim for the establishment of the RMCE system in *P. pastoris* was to develop a pilot system in order to provide a proof of principle for the functionality of the system in the yeast expression host. Therefore, a *P. pastoris* RMCE MCL should be generated through the insertion of an exchangeable RMCE cassette with all necessary genetic elements ("tagging").

The versatile multi-host vector pFlpBtM harbors all needed elements to serve as a donor vector for the RMCE (§ 1.4). Hence, the *P. pastoris* RMCE tagging cassette was designed to be compatible to pFlpBtM as a beneficial, straightforward way to also expand to the multi host expression system (mHost-XS) to *P. pastoris*. The exchange cassette of pFlpBtM relies on the presence of a stable, "endogenous" promoter upstream of the RMCE cassette to drive the expression of a gene of interest (GOI). The *GAP* promoter was selected to fulfill this task in *P. pastoris*, because it provides strong, constitutive gene expression. Concomitantly, it was decided to direct the RMCE tagging cassette to the genomic *GAP* locus by homologous recombination in the pilot phase to ensure comparability of the results without the risk of locus-dependent influences.

The principal setup of the envisioned RMCE cassette is shown in Figure 3.1. The exchangeable cassette flanked by the FRT sites F_3 and F_{WT} resides downstream of the *GAP* promoter. The cassette comprises the gene encoding the red fluorescent protein RudolphRFP as a model GOI to track the expression level of a generated *P. pastoris* RMCE MCL. Moreover, the *Sh ble* selection

marker was inserted into the tagging cassette to enable stringent primary selection of the transformed clones. Downstream of the RMCE tagging cassette, an inactive selection gene ("selection trap") is located. The selection trap lacks both an ATG start codon and a promoter in order to suppress its expression. The choice of a specific selection trap gene for *P. pastoris* was based on the fact that the yeast beneficially offers the use of amino acid biosynthesis genes as auxotrophic selection markers in combination with a suitable mutant strain. The application of an auxotrophic selection trap would enable a cost-efficient alternative selection method for exchanged clones in comparison to the cost-intensive antibiotic selection traps used in eukaryotic cell culture. Therefore, the *His4* gene was selected to serve as a pioneer model for the establishment of an auxotrophic selection trap for *P. pastoris* (Δ*his4*). Consequently, the histidine auxotroph *P. pastoris* strain GS115 (*his4*) was selected to generate the model RMCE MCL.

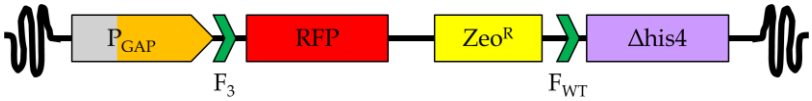


Figure 3.1. Schematic view of the envisioned RMCE tagging cassette in *P. pastoris* GS115.
The RudolphRFP marker gene (RFP) and the Zeocin antibiotic selection gene *Sh ble* (Zeo^R) are flanked by FRT F_3 and F_{WT}. P_{GAP}: *P. pastoris GAP* promoter to drive expression of any model protein or exchanged GOI. Δ*his4*: *His4*-based selection trap without ATG start codon and promoter. The transcription terminators of RFP and Δ*his4* as well as the promoter and the transcriptional terminator of Zeo^R were omitted from graphical display for simplifying reasons.

3.1.2 Construction of the *P. pastoris* RMCE tagging vector and generation of the pilot RMCE master cell line

For the generation of the *P. pastoris* RMCE MCL, the tagging vector was cloned with the desired features (§ 3.1.1), as described in detail in chapter 2.3. Briefly, the vector pGAPZα-A (Thermo Fisher Scientific) was chosen as a basis for the tagging vector, as it harbors the genomic *GAP* promoter. In the first step, the secretion signal sequence (MF-α ss) present on pGAPZα-A was deleted to allow for intracellular expression. Compared to pGAPZ-A (Thermo Fisher Scientific), the modified plasmid lacked an 8 bp spacer sequence between the *Bst*BI and *Eco*RI restriction sites of the multi cloning site ("Δ8"). In addition, an Ampicillin resistance cassette was inserted into the vector for cost-efficient selection in *E. coli*. The resulting plasmid was designated as pGAPZAΔ8.

In the next steps, the RMCE cassette and the Δ*his4* gene were inserted into pGAPZAΔ8 downstream of the *GAP* promoter. The 3' border of the RMCE cassette was cloned to harbor a complemented Δ*his4* gene ("ATG start codon – FRT F_{WT} – Δhis4") downstream of the TEF1/EM7 promoter, replacing the Zeocin resistance gene (*Sh ble*) on the vector. The resulting vector was named pYEXs ("yeast exchange simulation"), because it harbors the sequence of a RMCE cassette with a complemented Δ*his4* as it should look like after a cassette exchange. The vector pYEXs was used to clone a set of variant vector constructs to test the functionality and selectivity of Δ*his4*

under the control of different foreign promoters. The results of this test are presented in chapter 3.1.3. The tagging vector was generated in two additional cloning steps from pYEXs. First, the Zeocin resistance cassette was inserted back into the vector, replacing the "TEF1/EM7 promoter – ATG" combination to render Δ*his4* inactive again. The vector product was named pYTA ("yeast tagging"). Secondly, the *RudolphRFP* gene was incorporated into the MCS of pYTA downstream of the *GAP* promoter and FRT F_3. The resulting tagging vector was designated as pYTA-RudolphRFP (Figure 3.2).

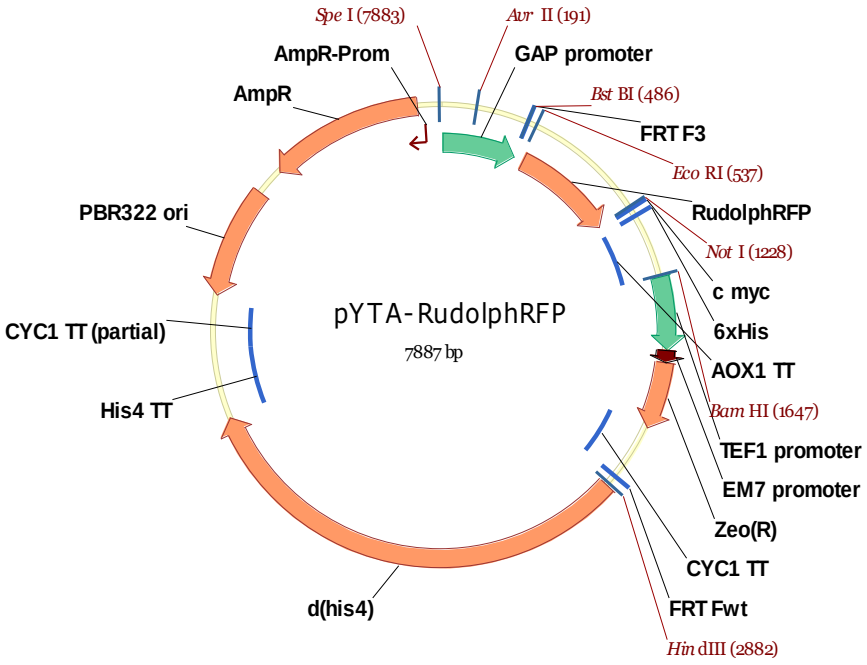


Figure 3.2. Vector map of pYTA-RudolphRFP.
The vector comprises a pBR322 ori, an ampicillin resistance cassette (AmpR) for bacterial selection. The RMCE tagging cassette flanked by FRT F_3 and FRT F_{WT} is located downstream of the *GAP* promoter. It includes *RudolphRFP* as a fluorescent marker gene and a Zeocin resistance (ZeoR) cassette for bacterial and yeast selection. Downstream of the RMCE cassette, the inactive selection trap Δ*his4* is located (d(his4)). The image was created with the software Vector NTI Suite 8 (Thermo Fisher Scientific). Cyan arrows: eukaryotic promoters. Dark red arrows: prokaryotic promoters. Orange arrows: coding sequences and origin of replication (ori). Blue bars: transcriptional terminator sequences (TT).

The tagging vector pYTA-RudolphRFP was linearized inside the *GAP* promoter and stably transformed into *P. pastoris* GS115 by electroporation to generate the RMCE MCL (§ 2.6.9). Low amounts of DNA (~25 ng) were used to avoid multi copy vector insertions. The transformed cells were selected on YPDS Zeocin agar plates. Eight single clones were isolated and tested by two PCR reactions from genomic DNA (§ 2.6.11). For simplification, clones will be designated according to the strain name and the vector (e. g. GS115/pYTA-RudolphRFP) from here on.

The first PCR (PCR1) was performed to validate the presence of the integrated vector through the amplification of a fragment of the vector containing FRT F_3 and *RudolphRFP*. The second PCR (PCR2) was conducted to differentiate between single copy clones and multi copy clones. Herein, a fragment was amplified, which would only be formed in case of tandem insertions of the vector

into the genome. This was done to ensure comparability of the different clones and to prevent the risk of undefined RMCE reactions due the presence of recombinable *cis* elements in case of tandem copies. A schematic depiction of the PCR strategy and the results are shown in Figure 3.3. The used primers and the expected fragment sizes are listed in Appendix II. The majority of the clones (apart from clone 5) were identified as single insertion clones.

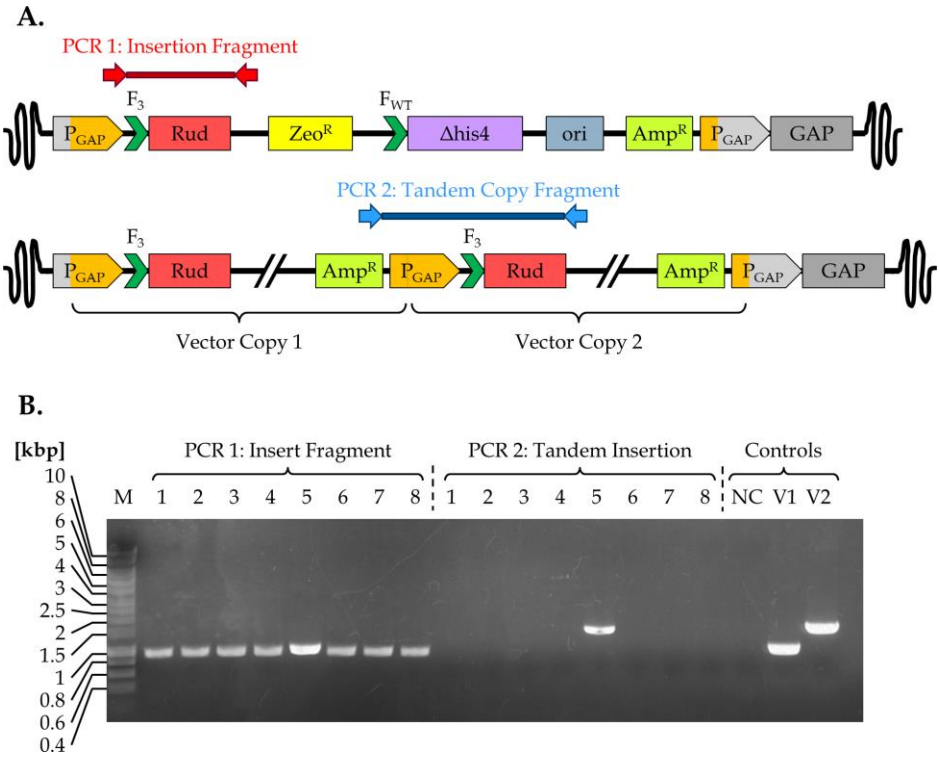


Figure 3.3. Validation of *P. pastoris* RMCE MCL by genomic PCR.
(A) The vector insertion is illustrated by the orange and grey colors of the *GAP* promoter from the vector and the genome, respectively. The PCR product to test the insertion of the vector into the genome is shown in red and spans over the gene of interest. A second PCR product (blue) can only occur in case of tandem insertions of the vector into the genome, as depicted for two inserted copies in the scheme. This allows differentiating between single copy and multi copy insertions. **(B)** The PCR products were loaded onto a 0.8% agarose gel, which was stained with ethidium bromide following gel electrophoresis. The bands indicate that all clones apart from clone 5 harbor single copy insertions of the vector. Numbers 1-8 indicate clone numbers. NC is a negative control of PCR 1 with genomic DNA of untransformed *P. pastoris* GS115. V1 and V2 are PCRs 1 and 2 with vector DNA of pYTA-RudolphRFP.

3.1.3 Evaluation of the novel selection trap Δ*his4* in *P. pastoris*

The *His4* gene was chosen as a model for a novel amino acid biosynthesis based / auxotrophic selection trap (Δ*his4*) in *P. pastoris*. The complementation of Δ*his4* following a RMCE reaction should restore the prototrophy of the *P. pastoris* strain GS115 (*his4*), enabling it to grow on histidine drop-out minimal medium. To determine the applicability of Δ*his4*, two major key points had to be assessed: First, its selectivity in both complemented and non-complemented states and secondly, its expression under the control of the foreign promoters from putative RMCE donor vectors. The growth of *P. pastoris* cell lines harboring a complemented Δ*his4* gene under the

control of different promoters was analyzed in comparison to the RMCE MCL with the supposedly inactive selection trap on selective minimal medium.

For this, a set of vectors was cloned from pYEXs (§ 3.1.2), which harbors the complemented Δ*his4* gene under the control of the TEF1/EM7 promoter, as described in chapter 2.3. The TEF1/EM7 promoter was known to be functional in *P. pastoris*. Hence, the main purpose of pYEXs-yeGFP was to evaluate the functionality of the complemented Δ*his4*. Additionally, two heterologous promoters were analyzed for their capability to induce the expression of the complemented Δ*his4* in *P. pastoris*. The first promoter was the human PGK1 (hPGK) promoter, present on the envisioned RMCE donor vector, pFlpBtM. The second promoter was the OpIE1 promoter from *Orgyia pseudotsugata*, which was a possible candidate to succeed the hPGK promoter used in the exchange cassette of pFlpBtM. In a first step to generate the respective vectors, the *y-eGFP* gene was cloned into pYEXs as a model GOI. This was done to additionally evaluate the intensity of the fluorescence of y-eGFP in comparison to RudolphRFP in this experiment. The resulting vector was named pYEXs-yeGFP. The two heterologous promoter candidates were then respectively cloned into pYEXs-yeGFP, replacing the TEF1/EM7 promoter. The vectors were designated as pYEXsPGK-yeGFP and pYEXsOpi-yeGFP. The vector maps are depicted in Appendix IV.1.

The three pYEXs variants (pYEXs-yeGFP, pYEXsPGK-yeGFP and pYEXsOpi-yeGFP) were linearized and stably transformed into *P. pastoris* GS115 (§ 2.6.10). Primary selection of the transformed clones was performed on minimal dextrose medium (MD) plates without supplemented histidine. Following single clone isolation, the clones were validated for single vector copy insertion by genomic PCR (§ 2.6.11). As described in chapter 3.1.2, two PCR reactions were performed per sample to validate the genomic insertion of the vector and to identify single copy clones. The first PCR comprised a fragment of the promoter upstream of the selection trap and FRT F_{WT}. The second PCR was used to amplify a product, which would only be formed in case of a tandem vector copies. A scheme of the PCR strategy and the PCR results are shown in Figure 3.4. The expected band sizes and the used primers are shown in Appendix II.

For each GS115/pYEXs variant, the single clones 3 and 6 were selected to perform the growth experiment. Additionally, the RMCE MCL clones GS115/pYTA-RudolphRFP 1 and 2 (§ 3.1.2) were also included in this experiment to assess the selectivity of the inactive Δ*his4*. All clones were serially diluted in 1 M sorbitol solution from freshly grown colonies and dropped onto MD (minimal dextrose) and MDH (MD + histidine) agar plates, as described in chapter 2.6.12. The growth of the clones after 48 h is depicted on Figure 3.5. The RMCE MCL could not grow on MD, but only on MDH. In contrast, all GS115/pYEXs variant clones were able to grow on MD at comparable levels. Based on these results, it could be concluded that Δ*his4* is feasible as a selection trap in *P. pastoris*. It is inactive in the RMCE MCL and its functionality is only restored by

complementation. Furthermore, the "complementingly restored" His4 protein was expressed at selective levels under the control of all three tested promoters with visually comparable levels of growth. This shows that the heterologous hPGK and OpIE1 promoters do not present an obstacle for the use of pFlpBtM as a RMCE donor vector in *P. pastoris*.

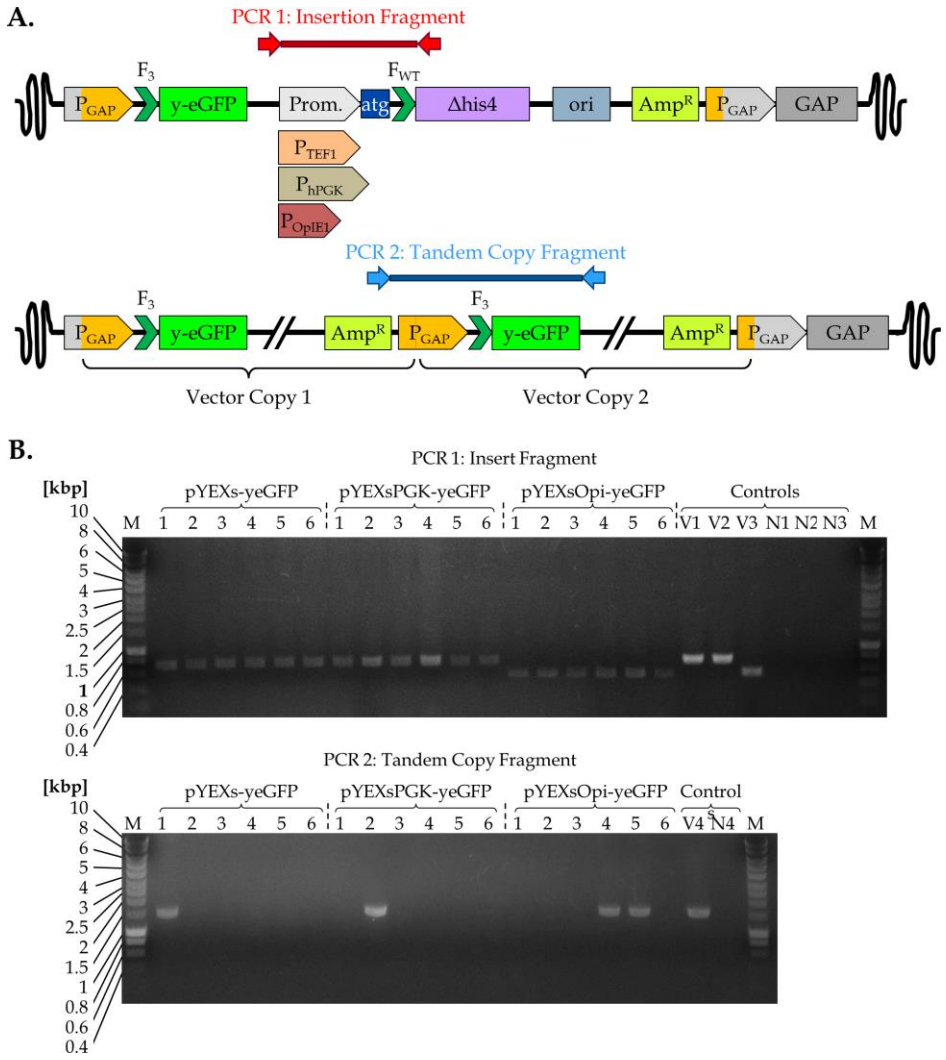


Figure 3.4. Validation of *P. pastoris* RMCE exchange simulation cell lines by genomic PCR.
(A) The vector insertion is illustrated by the orange and grey colors of the *GAP* promoter from the vector and the genome, respectively. The PCR product to assess the insertion of the vector into the genome is shown in red. It covers the respective promoter region. A second PCR product (marked in blue) will only be formed, if the vector is inserted multiple times in tandem copies, as depicted for two inserted copies in the picture. Single insertion clones were desired to ensure comparability of the different variants. **(B)** The PCR products were loaded onto a 0.8% agarose gel, which was stained with ethidium bromide following gel electrophoresis. The digits indicate the clone numbers. Vector names indicate the stably transformed plasmid. Negative control reactions of PCR1 and PCR2 were performed with unmodified GS115 cells (N1-N4), while the positive controls were amplified from the respective vector DNA (V1-V4). Most of the analyzed clones were single insertion clones (no signal in PCR2), which were usable for further analysis.

Finally, this experiment shows that the proposed concept of auxotrophic selection traps is applicable in *P. pastoris*. The feasibility of Δ*his4* delivers a proof of principle. It is conceivable that the results obtained for Δ*his4* can be transferred to other amino acid biosynthesis genes using respective auxotrophic mutant strains (e. g. *arg4*). Furthermore, the concept should be generally

usable independently of the RMCE system. All in all, the use of auxotrophic selection traps presents a novel, cost-efficient alternative to antibiotic selection traps in *P. pastoris*.

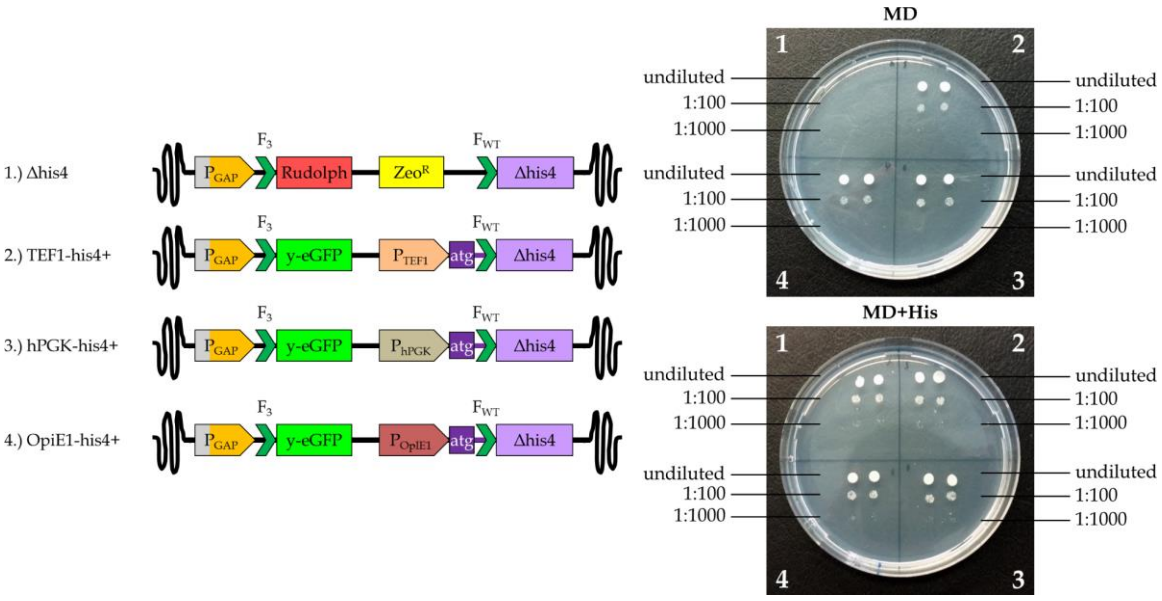


Figure 3.5. Evaluation of selection trap Δ*his4* by growth on selective minimal medium.
The scheme on the left-hand side shows the schematic cassettes of each variant. The experiment was performed with clones 1 and 2 of the RMCE MCL (1) as well as clones 3 and 6 of each exchange simulation variant (2-4). All tested clones were identified as single insertion clones. The cells were respectively dropped onto agar plates with minimal dextrose medium (MD) and MD supplemented with histidine (MD+His) in three different dilutions. Growth was documented after 48 h at 30 °C (right-hand side of the figure). The RMCE MCL (1) with the putatively inactive Δ*his4* gene could not grow on MD. In contrast, all exchange simulation cell lines (2-4) with the complemented Δ*his4* (His4+) could grow on MD. The complementation of Δ*his4* under the control of all three tested promoters (TEF1, hPGK and OpIE1) resulted in similar levels of growth. All clones could grow on the control agar plate (MD+His).

3.1.4 Evaluation of marker protein expression in the *P. pastoris* pilot RMCE master cell line

During the selection trap analysis (§ 3.1.3) neither the pilot RMCE MCL nor the GS115/pYEXs variant clones visibly expressed RudolphRFP or y-eGFP under the control of the constitutive *GAP* promoter. Hence, the expression was analyzed and the role of several structural elements causing this possible bottleneck in expression was evaluated.

3.1.4.1 Investigation of the influence of structural elements of the RMCE cassette on gene expression in *P. pastoris*

The RMCE cassette of the pilot RMCE MCL was designed to be compatible to pFlpBtM as a donor vector. Hence, the FRT F_3 site is located between the "endogenous" *GAP* promoter and the model GOI (*RudolphRFP* or *y-eGFP*). In regard of the expression cassette of the GOI, the presence of FRT F_3 in this position comprises the major difference between the generated vectors pYTA/pYEXs and the commercially available pGAPZ vectors (Thermo Fisher Scientific). For this reason, it was speculated that FRT F_3 in its current location might negatively influence the *GAP*

promoter based expression of the fluorescent marker proteins in *P. pastoris*. However, it was not known, how strong this influence was or whether it was caused by an unknown, exclusive cross-interaction of FRT F_3 and the *GAP* promoter.

To evaluate this, the *AOX1* promoter was cloned into the tagging vector pYTA-RudolphRFP, replacing the *GAP* promoter. The *AOX1* promoter is the strongest available, native promoter in *P. pastoris*. If the expression of the combination "promoter – F_3 – GOI" is not restored by this promoter, then it is conceivable that no other promoter would be able to overcome this bottleneck in *P. pastoris*. In addition to the *AOX1* promoter, the modified tagging vector was designed to harbor the genomic 3'-*AOX1* region downstream of the selection trap. This was done to mediate a more stringent single insertion of the tagging vector via double crossing over and gene replacement. In consequence of this strategy, the vector linearization to transform *P. pastoris* was performed with *Bgl*II. Since *RudolphRFP* harbors an internal *Bgl*II restriction site, it was replaced with *mCherry* in the *AOX1* tagging vector. The final vector was named pYTAaox-mCherry. In addition to the *AOX1* tagging vector, a set of "classical" *GAP* and *AOX1* producer cell lines was generated for a comparative expression experiment of the fluorescent marker proteins RudolphRFP, y-eGFP and mCherry. The respective genes were cloned into the vectors pGAPZAΔ8 (§ 3.1.2) and pPICZ-A (Thermo Fisher Scientific). In total, six vectors were cloned: pPICZ-RudolphRFP, pPICZ-y-eGFP, pPICZ-mCherry, pGAPZAΔ8-RudolphRFP, pGAPZAΔ8-yeGFP and pGAPZAΔ8-mCherry. The detailed cloning strategy of all generated vectors is described in chapter 2.3. The vector maps are depicted in Appendix IV.2. The six test vectors and the *AOX1* tagging vector were respectively linearized and transformed into *P. pastoris* GS115 as described in chapter 2.6.10. Positive clones for each transformed construct were selected on YPDS Zeocin agar plates. After single clone isolation, single copy insertion of the vector was evaluated by genomic PCR (§ 2.6.11) in two reactions, as described before in chapter 3.1.2. The first PCR (PCR1) was conducted to validate the stable integration of the vector. The second PCR (PCR2) was performed to identify single copy insertion clones. The expected band sizes of the PCR products and the used primers are listed in Appendix II. The results are depicted in Figure 3.6, Figure 3.7 and Figure 3.8.

Two single insertion clones per vector transformation were isolated to evaluate the expression of the fluorescent proteins under the control of the *AOX1* and the *GAP* promoter with and without FRT F_3 residing between promoter and GOI. For all generated cell lines, the clones 1 and 2 were selected – except for GS115/pPICZ-RudolphRFP, for which the clones 2 and 3 were used. The selected clones were incubated in 100 mL shake flask cultures according to chapter 2.8.1. The previously generated *GAP* RMCE MCL was additionally included in the analysis (GS115/pYTA-RudolphRFP, clones 1 and 2, § 3.1.2). For the expression of the fluorescent markers under the control of the *GAP* promoter, the clones were cultivated in BMDY medium (glucose as a carbon

source). For the *AOX1* promoter driven expression of the marker proteins, the clones were first grown in BMGY medium (glycerol) and then transferred to BMMY medium to induce the expression (methanol as a carbon source). After 48 h of expression, samples of all cultures were taken and analyzed under UV light (330 nm) for visible fluorescence (Figure 3.9).

Significantly visible expression was obtained from all six fluorochrome test cell lines without FRT F_3. In contrast, both RMCE MCLs (*AOX1* and *GAP*) did not show any expression. Even the strong *AOX1* promoter did not restore the expression to visible levels. This shows that the bottleneck in expression cannot be overcome in this way. Moreover, in regard of the expression cassette "Promoter – GOI – Terminator", the presence of FRT F_3 between promoter and GOI comprises the major difference between the test cell lines and the RMCE MCLs. The expression was significantly hampered for both promoters only in the presence of FRT F_3. This strongly supports the hypothesis that FRT F_3 is the critical genetic element causing the negative impact on the expression of the fluorescent marker proteins (RudolphRFP and mCherry) in *P. pastoris*.

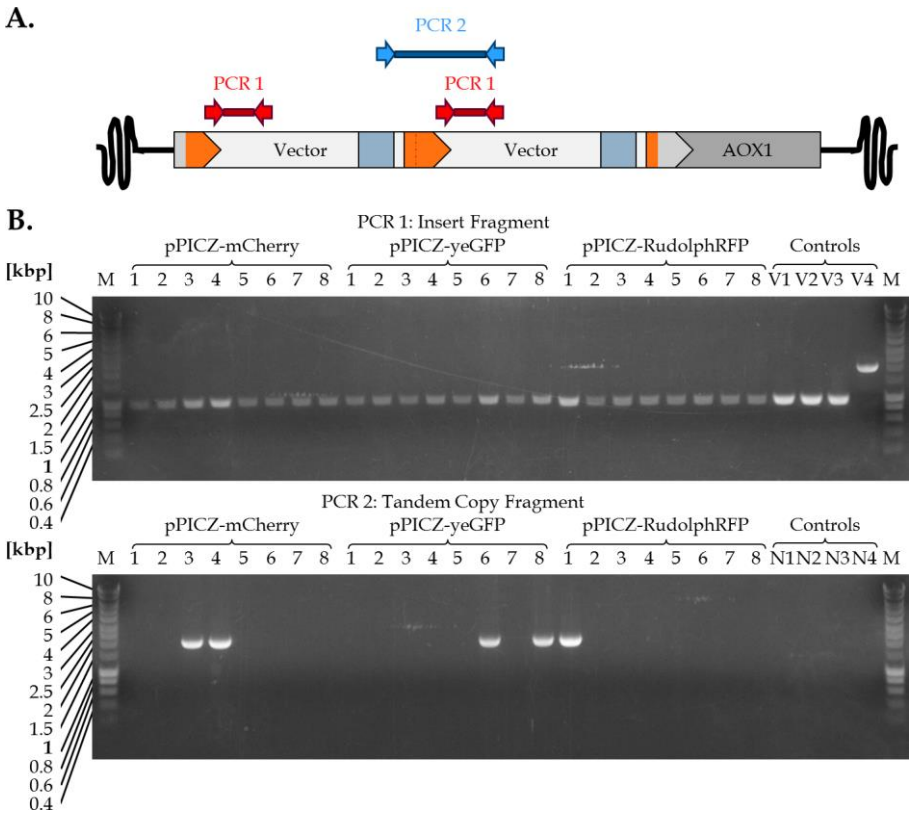


Figure 3.6. Validation of *P. pastoris* fluorochrome expression test cell lines by genomic PCR.
(A) Simplified scheme showing the PCR products of the stably inserted fluorochrome test vectors based on pPICZ-A. Orange/grey arrow: *AOX1* promoter (bicolored to indicate the homologous recombination); blue box: origin of replication (ori); AOX1 (dark grey): genomic *AOX1* gene. Two PCRs were performed to evaluate the clones. PCR 1 (red) comprises the GOI to validate the insertion of the vector into the genome. PCR2 (blue) will only be yield a product in case of tandem insertions, as depicted for two inserted vector copies in the figure. **(B)** Results of the electrophoresis of the PCR products on 0.8% agarose gels, which were stained with ethidium bromide. The vector names denote the respective plasmid used to transform *P. pastoris* GS115. The digits indicate the clone numbers. Negative control reactions were performed with unmodified GS115 cells (PCR1: N1-N3 / PCR2: N4). All clones showing a signal in PCR2 present multi copy clones and were not used in the further analysis. The majority of the clones present single copy insertions.

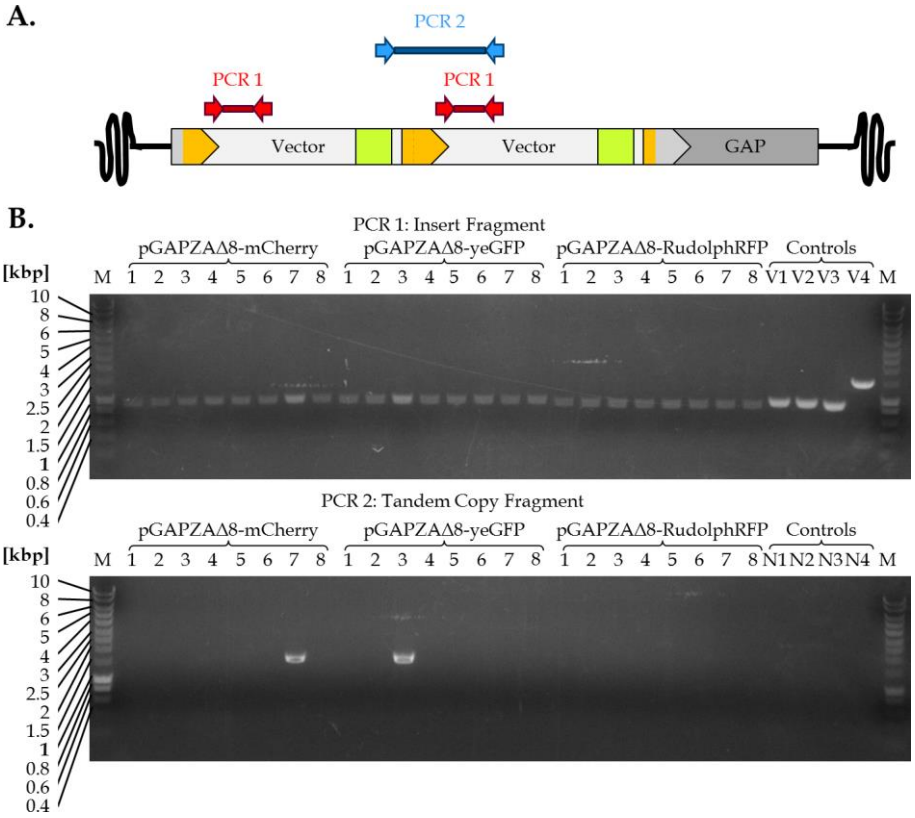


Figure 3.7. Validation of *P. pastoris* GAP fluorochrome expression test cell lines by genomic PCR.
(A) Simplified scheme showing the PCR products of the stably inserted fluorochrome test vectors based on pGAPZAΔ8. Light orange/grey arrow: *GAP* promoter (bicolored to indicate the homologous recombination); blue box: origin of replication (ori); GAP (dark grey): genomic *GAP* gene. Two PCRs were performed to evaluate the clones. PCR 1 (red) comprises the GOI to validate the insertion of the vector into the genome. PCR2 (blue) will only be yield a product in case of tandem insertions, as depicted for two inserted vector copies in the figure. **(B)** Results of the electrophoresis of the PCR products on 0.8% agarose gels, which were stained with ethidium bromide. The vector names denote the respective plasmid used to transform *P. pastoris* GS115. The digits indicate the clone numbers. Negative control reactions were performed with unmodified GS115 cells (PCR1: N1-N3 / PCR2: N4). All clones showing a signal in PCR2 present multi copy clones and were not used in the further analysis. The majority of the clones present single copy insertions.

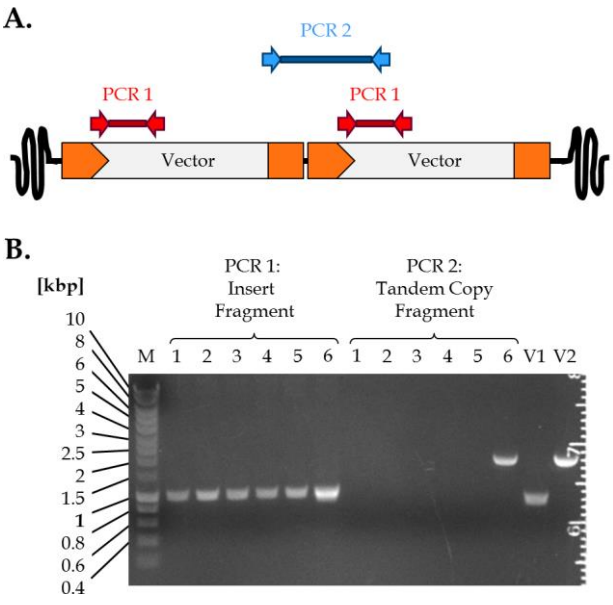


Figure 3.8. Validation of the *P. pastoris AOX1* RMCE MCL by genomic PCR.
(A) Simplified scheme showing the PCR products of the stably inserted tagging vector pYTAaox-mCherry. Orange/grey arrow: *AOX1* promoter; orange box: 3'-AOX1 region. Two PCRs were performed to evaluate the clones. PCR 1 (red) comprises the GOI to validate the insertion of the vector into the genome. PCR2 (blue) will only be yield a product in case of tandem insertions, as depicted for two inserted vector copies in the figure. **(B)** Results of the electrophoresis of the PCR products on 0.8% agarose gels, which were stained with ethidium bromide. The vector names denote the respective plasmid used to transform *P. pastoris* GS115. The digits indicate the clone numbers. Since the PCR reactions were performed in parallel to the ones used for the fluorochrome test cell lines, the control reactions are covered by samples N1-N4 in Figure 3.7. C1-C5 present single insertion clones showing only a signal in PCR1.

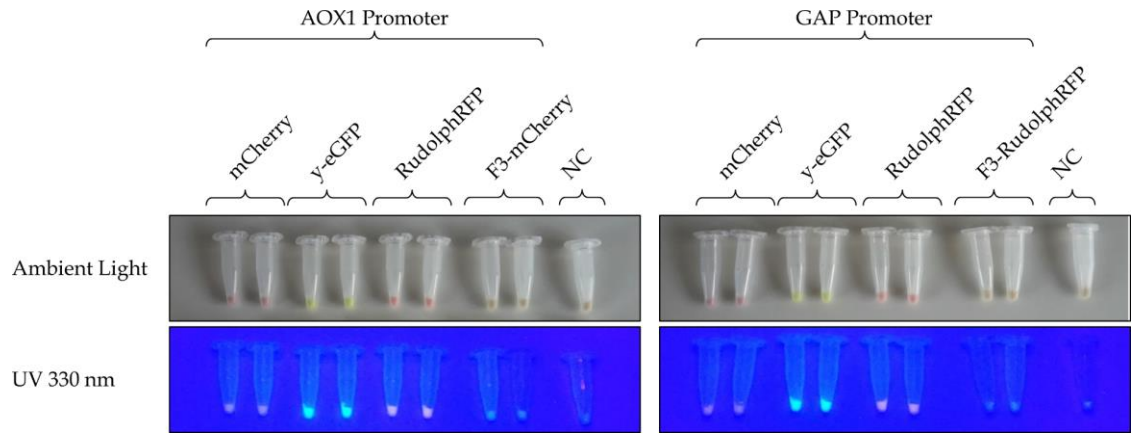


Figure 3.9. Expression of mCherry, y-eGFP and RudolphRFP under the control of the *AOX1* promoter and the *GAP* promoter from *P. pastoris* GS115 transformed with vectors with or without FRT F_3.
Samples were documented after 48 h of expression under ambient light and UV light (330 nm). The fluorochrome test cell lines without FRT F_3 are depicted by the name of the transformed fluorescent marker (mCherry, y-eGFP, RudolphRFP). The *AOX1* RMCE MCL is depicted as F3-mCherry, the *GAP* RMCE MCL as F3-RudolphRFP. Two single insertion clones of each cell line were analyzed. The groups "*AOX1* Promoter" and "GAP Promoter" indicate the clones expressing the GOI respectively under the control of the *AOX1* or the *GAP* promoter. Empty GS115 cells served as a negative control (NC). All test cell lines without FRT F_3 show visible fluorescence. In contrast, both RMCE MCL clones only display auto fluorescence (also visible for the NC), but do not show any expression of the respective RFP marker protein.

3.1.4.2 Insertion of the *GAP* promoter inside of the *AOX1* RMCE tagging cassette

It was observed that the expression of fluorescent marker genes is hindered in the presence of FRT F_3 between promoter and GOI. To initially test whether a different position of FRT F_3 would provide a simple way to restore gene expression, the *GAP* promoter was inserted downstream of FRT F_3 as part of the exchangeable cassette of the *AOX1* tagging vector pYTAaox-mCherry. The resulting expression cassette of *mCherry* ("Promoter – *mCherry* – Terminator") would be virtually identical to the vector pGAPZAΔ8-mCherry, which had led to visible marker expression in previous experiment (§ 3.1.4.1). If this modification would already restore expression, it would even conveniently allow using the existing *AOX1* RMCE MCL (§ 3.1.4.1) with a modified RMCE donor vector to perform the cassette exchange.

The *GAP* promoter was inserted into pYTAaox-mCherry between FRT F_3 and *mCherry* (§ 2.3). The vector product was named pYTAaox-GAP-mCherry. The vector maps are depicted in Appendix IV.3. The vector was transformed into *P. pastoris* GS115 in two different ways. First, the vector was linearized inside the *AOX1* promoter and the 3'*AOX1* region to insert it into the genomic *AOX1* locus. Thereby, a RMCE MCL was generated, which should lead to the expression of mCherry under the control of the GAP promoter. The resulting clones were named GS115/pYTAaox-GAP-mCherry$^{(AOX1)}$. Secondly, as an additional control, the vector was linearized inside the *GAP* promoter to direct it to the genomic *GAP* locus instead of the *AOX1* locus. Thereby, the combination of *AOX1* promoter and FRT F_3 was placed upstream of the genomic *GAP* gene. The resulting clones were named GS115/pYTAaox-GAP-mCherry$^{(GAP)}$. A schematic overview of these insertions is presented in Figure 3.10A.

Primary selection and single clone isolation were performed on YPDS Zeocin plates. Following the primary selection, the control integration of the vector into the *GAP* locus (GS115/pYTAaox-GAP-mCherry$^{(GAP)}$) resulted in only five grown colonies with reduced growth rates. In contrast, a significantly higher amount of grown colonies was observed for GS115/pYTAaox-GAP-mCherry$^{(AOX1)}$. The corresponding transformation plates after primary selection are depicted in Figure 3.10B. Single clones were isolated for GS115/pYTAaox-GAP-mCherry$^{(AOX1)}$ and GS115/pYTAaox-GAP-mCherry$^{(GAP)}$.

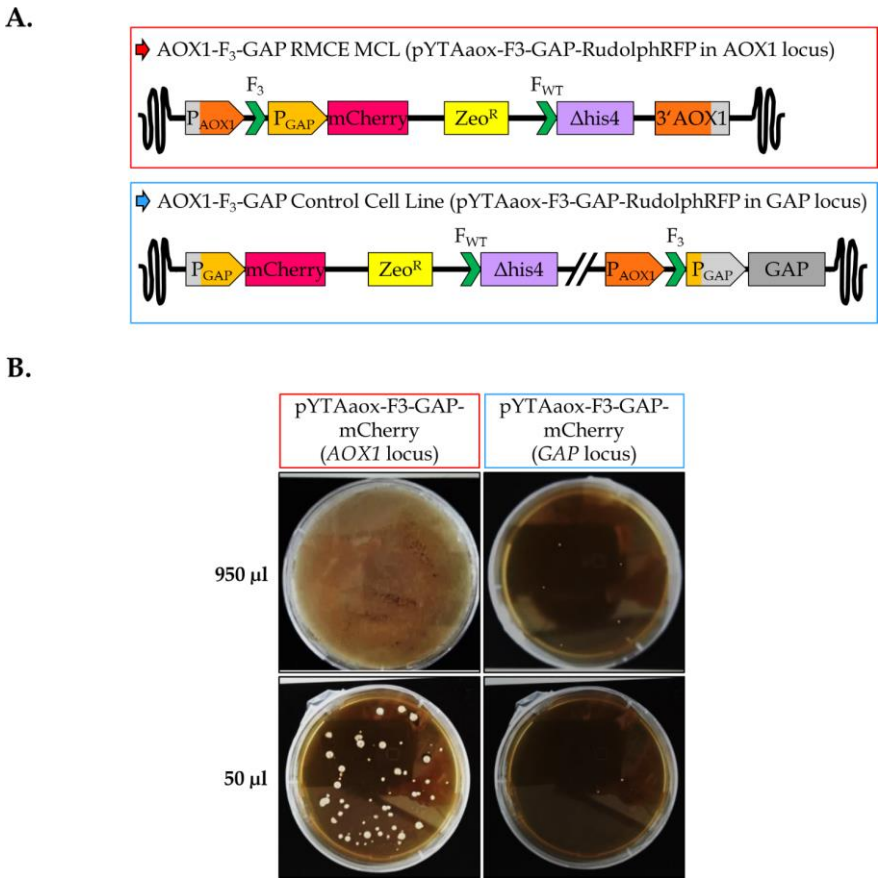


Figure 3.10. Overview of the respective transformations of pYTAaox-GAP-mCherry into *P. pastoris* GS115. **(A)** The scheme of pYTAaox-GAP-mCherry integrating into the genomic *AOX1* locus (red; *AOX1*-F3-*GAP* RMCE MCL) or the *GAP* locus (blue; control cell line) is shown. **(B)** YPDS Zeocin agar plates of the transformed GS115 clones after growth for 72 h of growth at 30 °C. It is shown that the insertion of the vector into the *AOX1* locus led to significantly higher amounts of colony forming units compared to the insertion into the *GAP* locus. The RFP mCherry is not expressed at visible levels.

The purified clones were validated by genomic PCR (§ 2.6.11). Two reactions per clone were performed to verify the integration of the vector into the genome and to identify single insertion clones, as previously described (§ 3.1.2). A schematic overview and the results are shown in Figure 3.12 (A and B). The used primers and the expected band sizes are listed in Appendix II. All clones of GS115/pYTAaox-GAP-mCherry$^{(AOX1)}$ were shown to be single insertion clones. In contrast, the five clones of GS115/pYTAaox-GAP-mCherry$^{(GAP)}$ were proven to be false-positives by the genomic PCR analysis. The isolated clones were respectively streaked out again on fresh YPD agar

plates to assess the expression of mCherry in comparison to the previously tested clones 3 and 6 of GS115/pGAPZAΔ8-mCherry (§ 3.1.4.1). The plates were documented after 72 h of incubation (Figure 3.11C). No expression of mCherry was observed from GS115/pYTAaox-GAP-mCherry$^{(AOX1)}$ or the false-positive clones of GS115/pYTAaox-GAP-mCherry$^{(GAP)}$. However, mCherry was visibly expressed from GS115/pGAPZAΔ8-mCherry.

Taken together, the results show that the expression of mCherry was not restored by placing the *GAP* promoter inside the *AOX1* tagging cassette. The "*AOX1* promoter – FRT F_3" combination upstream of the *GAP* promoter is interfering with the activity of the downstream *GAP* promoter. This was deduced for two reasons. First, the integration of the vector into the *AOX1* locus resulted in a lack of fluorescent cells. The substantial difference between pYTAaox-GAP-mCherry and the pGAPZAΔ8 derived fluorochrome test vectors (§ 3.1.4.1) regarding the expression cassette "*GAP* promoter – mCherry – Terminator" is the presence of the upstream "*AOX1* promoter – FRT F_3" combination. Secondly, the control insertion into the *GAP* locus did not yield any positive clones. Directed insertion should result in the "*AOX1* promoter – FRT F_3" combination residing upstream of the endogenous *GAP* gene (Figure 3.10A). If the *GAP* promoter is indeed knocked out by this combination, then this would result in a lack of glyceraldehyde-3-phosphate dehydrogenase (GAP). This would explain the notable absence of positive clones, since the GAP is an essential enzyme for cell maintenance and survival. However, the experiment left open which of the two elements (FRT F_3 site or *AOX1* promoter) was responsible for the lack of expression.

FRT F_3 had already been shown to lead to a lack of expression in a position between promoter and GOI (§ 3.1.4.1). Hence, it might still play a role in its current position upstream of the *GAP* promoter. The *AOX1* promoter could also be the responsible element, as it is tightly repressed under glucose conditions. This regulatory repression might have influenced the downstream residing *GAP* promoter. Finally, the lack of expression could have been caused by an undefinable influence through combined effects by overlapping sequences or spacing of the genetic elements. In the end, the results clearly show that the modified RMCE MCL GS115/pYTAaox-GAP-mCherry$^{(AOX1)}$ did not present a viable alternative for the establishment of the RMCE in *P. pastoris*.

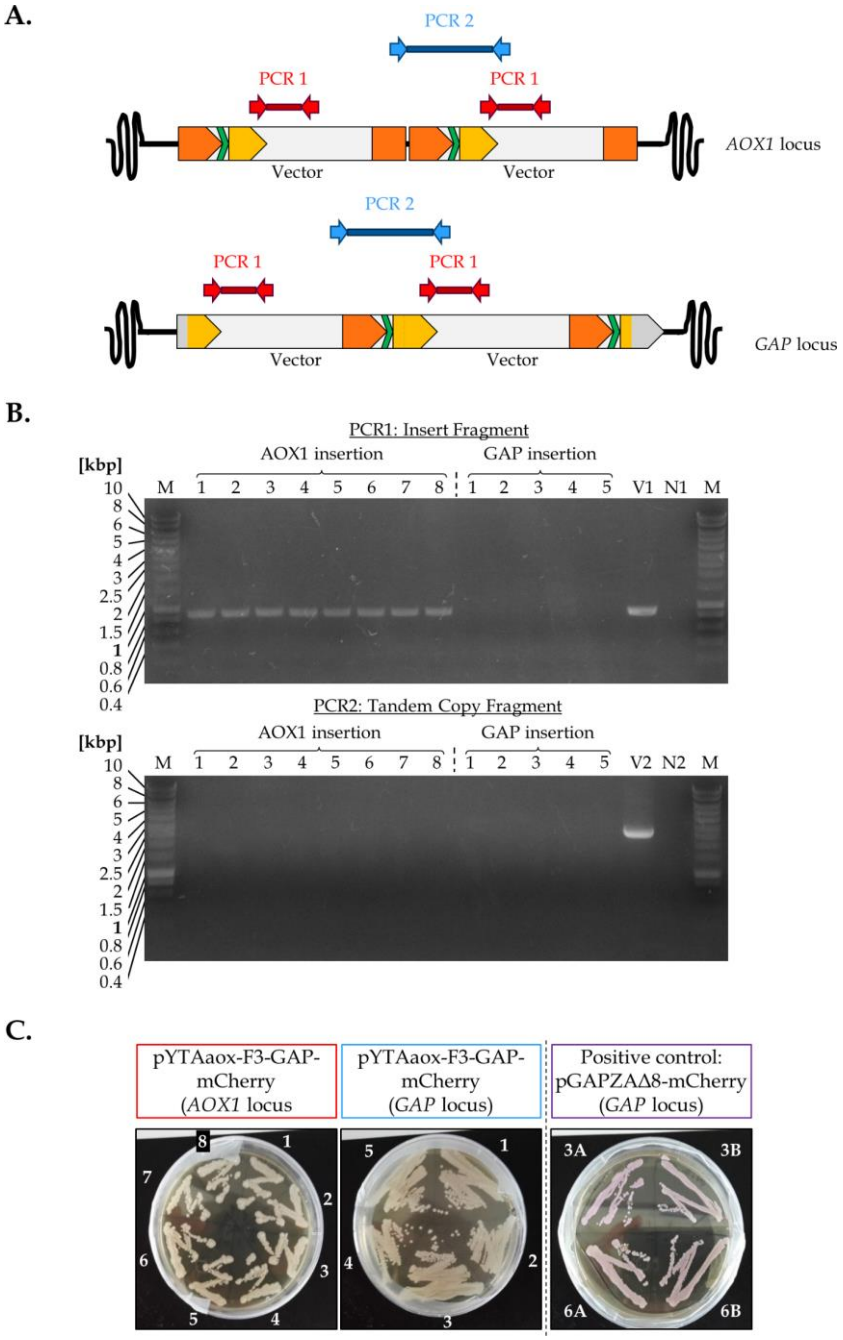


Figure 3.11. Analysis of the *AOX1*-F3-*GAP* RMCE MCL and the respective control insertion.
(A) Simplified scheme of the possible PCR products following the insertion of pYTAaox-GAP-mCherry into the *AOX1* or the *GAP* locus of *P. pastoris* GS115. Dark orange arrow: *AOX1* promoter; dark orange box: 3'*AOX1*; green arrow: FRT F_3; light orange/grey arrow: GAP promoter (bicolored depiction indicates homologous recombination). Two PCRs were performed to evaluate the clones. PCR 1 (red) comprises the GOI to validate the insertion of the vector into the genome. PCR2 (blue) will only yield a product in case of tandem insertions, as indicated for two vector copies in the picture. **(B)** Results of the electrophoresis of the PCR products on 0.8% agarose gels stained with ethidium bromide. The digits represent the clone numbers. The control reactions apply for both sets of clones. For N1 and N2, genomic DNA from untransformed GS115 cells was used as a negative control. For V1 and V2, the PCR was conducted with the vector as a positive control. None of the *AOX1*-F3-*GAP* RMCE MCL clones showed tandem insertions. All of the control clones (GAP insertion) were false positives, which did not show any PCR product. **(C)** Restreaks of the clones on YPD agar plates. Digits indicate the clone numbers. As a positive control, clones 3 and 6 of GS115/pGAPZAΔ8-mCherry (§ 3.1.4.1) were additionally incubated in two technical replicates (A, B). Only the positive control showed visible expression of mCherry after incubation at 30 °C for 72 h.

3.1.4.3 Design and generation of modified *P. pastoris* RMCE master cell lines

In the previous experiments, it was observed that the pilot RMCE MCL (GS115/pYTA-RudolphRFP) did not show any expression of RudolphRFP. FRT F_3 was identified to cause the lack of expression in a position between promoter and GOI (§ 3.1.4.1). However, the mechanism of this reduction in expression remained unclear. The bottleneck in expression could not be solved by placing the *GAP* promoter inside the tagging cassette of pYTAaox-mCherry (§ 3.1.4.2). Hence, new variants of the RMCE tagging cassette were designed and generated by repositioning FRT F_3. The experiment was conducted to evaluate three important key points.

1. It should be validated that FRT F_3 was the single, specific element inside the tagging vector to be responsible for the observed lack of expression.
2. Furthermore, it should be assessed, if the negative influence of FRT F_3 on the expression of the model GOIs occurs on the level of transcription or the level of translation.
3. Finally, a viable alternative to the pilot RMCE MCL with restored expression of the GOI should be identified in order to establish the RMCE system in in *P. pastoris*.

To do this, a new set of *P. pastoris* RMCE tagging vectors was designed and cloned to generate additional RMCE MCL variants. All vectors were designed to be directed to the genomic *GAP* locus by homologous recombination. In accordance, the *GAP* promoter was selected to drive the expression of the model GOI *RudolphRFP* in the modified RMCE tagging cassettes. This was done to ensure optimal comparability of the different RMCE MCL variants, while also excluding any influences by the genomic locus or genetic upstream elements (e. g. the *AOX1* promoter). In total, four modified variants of the initial tagging vector pYTA-RudolphRFP were generated. For simplification, pYTA-RudolphRFP will be referred to as variant V1 from here on. The detailed cloning strategy of the generated vectors is described in chapter 2.3. The respective vector maps are depicted in Appendix IV.4.

For variant V2, FRT F_3 was deleted from pYTA-RudolphRFP ("ΔF_3") to evaluate, whether FRT F_3 comprised the exclusively responsible genetic element for the observed lack of expression. The vector product was named pYTAΔF3-RudolphRFP. Variant V3 was based on the hypothesis that FRT F_3 in its original position between promoter and GOI would be transcribed to the 5'-UTR (untranslated region) of the mRNA. FRT sites comprise short inverted sequence repeats. It is conceivable that the transcription of F_3 results in the formation of a secondary structure in the 5'-UTR of the mRNA, which might negatively influence the protein translation. To evaluate this hypothesis, FRT F_3 was moved into the promoter region of the *GAP* promoter ("F_3-PR") on pYTA-RudolphRFP upstream of its TATA box. Thereby, FRT F_3 was eliminated from the transcribed 5'-UTR. The resulting tagging vector V3 was designated as pYTAF3PR-RudolphRFP.

Variant V4 was generated to mimic the sequence used to complement the selection trap Δ*his4*, which had resulted in the expression of His4 at selective levels (§ 3.1.3). Hence, the presence of FRT F_{WT} in the 5' region of the mRNA did apparently not cause a complete loss of expression, in contrast to what was observed for FRT F_3. *In silico* RNA secondary structures predictions with RNAFold (University of Vienna, Vienna, Austria) displayed similar minimum free energies for FRT F_3 (-16.06 kcal/mol) and FRT F_{WT} (-15.28 kcal/mol). Therefore, the influence of their secondary structures should be comparable. However, unlike F_3, F_{WT} is part of the open reading frame (ORF). F_3 on the other hand resides in the 5'-UTR upstream of the ATG start codon. This positional difference could possibly be critical for the expression of the respective gene. To evaluate this, the sequence preceding FRT F_{WT} in pFlpBtM (starting with the ATG start codon for the selection trap complementation) was inserted upstream of FRT F_3 inside pYTA-RudolphRFP ("ATG-F_3"). The vector product was named pYTAatgF3-RudolphRFP.

In the final variant V5, the *GAP* promoter was repositioned downstream of FRT F_3 inside the RMCE cassette. This setup is "essentially identical" to the above-described vector pYTAaox-GAP-mCherry (§ 3.1.4.2), but without the risk of any influence caused by the upstream *AOX1* promoter. To ensure optimal comparability to the other variants (V1-V4), tagging vector V5 should also be directed to the genomic *GAP* locus. However, in variant V5, the *GAP* promoter was part of the RMCE cassette. Consequently, it could not be used as a homologous sequence to direct the tagging vector to the *GAP* locus. This would compromise the RMCE cassette, similar to the previously described control insertion of pYTAaox-GAP-mCherry into the *GAP* locus (see Figure 3.10A, page 73, blue box). Hence, to enable the integration at the correct position, a suitable chromosomal upstream region of *GAP* ("URG") was needed to mediate the homologous recombination. The URG was selected from the chromosomal DNA sequence of *P. pastoris* GS115, comprising the first 508 bp upstream of the genomic *GAP* promoter (Genbank #FN392320, positions 809,079 – 809,598). Advantageously, the URG harbors a recognition site for *Sph*I, which was usable for vector linearization. To clone the tagging V5, the vector pYTA-RudolphRFP (V1) was modified in multiple cloning steps. Briefly, the combination "URG – F_3 – *GAP* promoter – *RudolphRFP*" was inserted into the vector in place of the "*GAP* promoter – F_3 – *RudolphRFP*" combination. The resulting vector was named pYTAUR-RudolphRFP.

The RMCE MCL V1 (GS115/pYTA-RudolphRFP; GS115/V1) had already been generated as described in chapter 3.1.2. To generate the new variants, the tagging vectors V2 to V5 were linearized and transformed into *P. pastoris* GS115. The empty tagging vector pYTA was also linearized and transformed into GS115 to create a negative control (NC) for the expression analysis. The existing clones of GS115/pGAPZAΔ8-RudolphRFP (§ 3.1.4.1) were selected as a positive control for the expression experiments (PC). A schematic overview of the seven cell lines (GS115/V1-V5, GS115/NC and GS115/PC) is depicted in Figure 3.12.

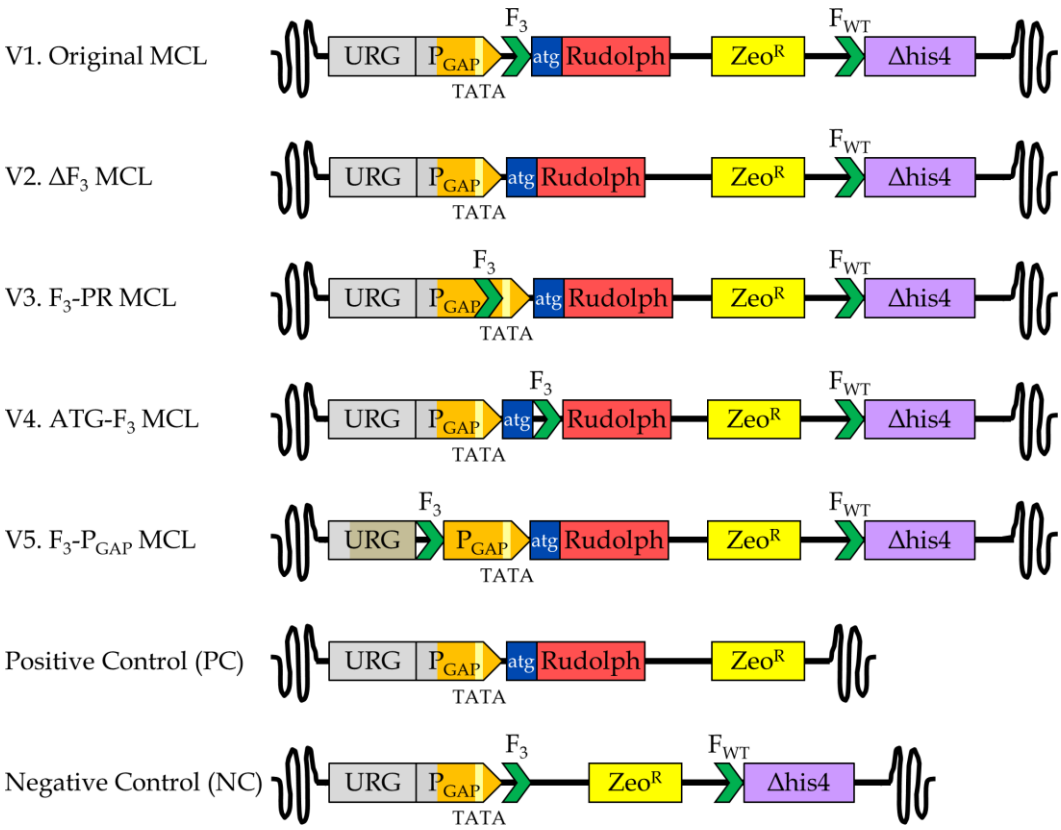


Figure 3.12. Schematic overview of the different *P. pastoris* RMCE MCL variants.
The figure depicts the genetic key elements of the RMCE MCL variants (V1 to V5) as well as the positive and negative control cell lines (PC and NC). The main difference between variants V1-V5 is the position or presence of FRT F_3, which flanks the 5′ border of the RMCE cassette. For simplification, the transcriptional terminator of *RudolphRFP* (Rudolph), the promoter and transcriptional terminator of the Zeocin resistance cassette (Zeo^R) and the vector elements downstream of the selection trap Δ*his4* have been omitted from graphical display. All vectors were inserted into the genomic *GAP* locus of *P. pastoris*. Therein, the *GAP* promoter (P_{GAP}) is located downstream of the genomic upstream region of *GAP* (URG). The ATG start codon of *RudolphRFP* and the TATA box of the *GAP* promoter are specifically denoted in the scheme, since they were subject to modifications in V3 and V4, respectively. Dual colors indicate the regions that were used to perform the homologous recombination (PC, NC and V1-V4: P_{GAP}; V5: URG).

Eight single colonies of each of the newly generated cell lines (GS115/V2-V5 and GS115/NK) were isolated and analyzed by genomic PCR (§ 2.6.11). Two PCRs were carried out per clone to verify the stable transformation of the vector and to identify single insertion clones. The first PCR product (PCR1) comprised a fragment spanning over the URG and *RudolphRFP*. The second PCR product (PCR2) could only be amplified in case of multi copy clones (as described in chapter 3.1.2). The forward primer in PCR1 (s_gGAP-F) binds to a genomic DNA sequence upstream of the URG. Using this primer, the PCR1 product will only be formed, if the vector is inserted in the correct genomic locus. For this reason, the previously tested eight respective clones of GS115-pYTA-RudolphRFP and GS115-pGAPZAΔ8-RudolphRFP were also included in the PCR analysis. The expected band sizes of the PCR products and the used primers are listed in Appendix II. The results of the PCRs are depicted in Figure 3.13.

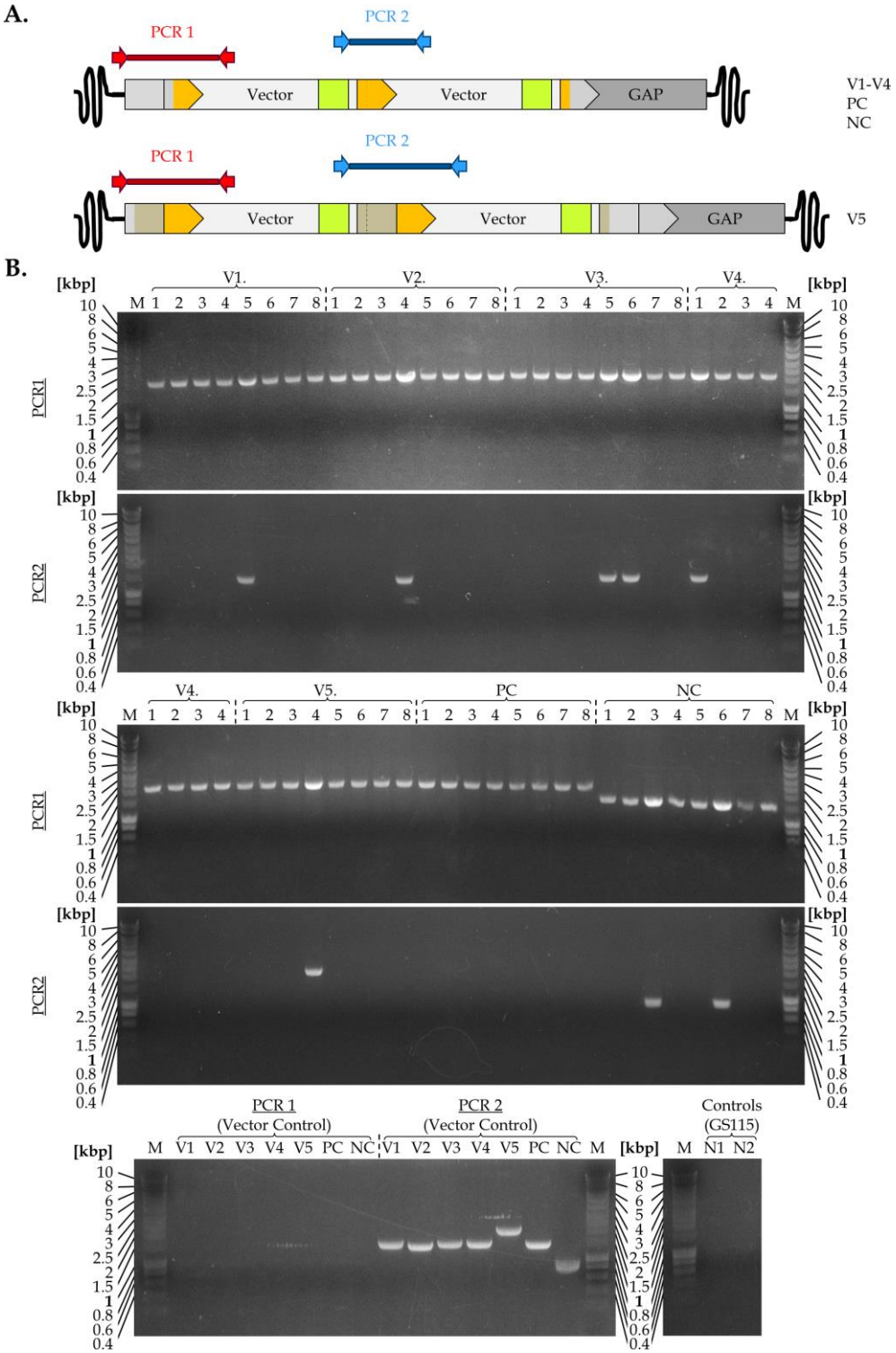


Figure 3.13: Validation of the new *P. pastoris* RMCE MCL variants by genomic PCR.
(A) Simplified scheme showing the PCR products of the vector insertion into the *GAP* locus of *P. pastoris*. Orange/grey arrow: *GAP* promoter; grey/ochre: URG (bicolored elements indicate the homologous recombination). Green: Amp resistance cassette; GAP (dark grey): genomic *GAP* gene. Two PCRs were performed to evaluate the clones. PCR 1 (red) comprises the GOI to validate the insertion of the vector into the genome. PCR2 (blue) will only yield a product for multi copy clones (depicted for two inserted vector copies in the picture). Single insertion clones were desired to ensure comparability between different clones. **(B)** Results of the electrophoresis of the PCR products on 0.8% agarose gels, which were stained with Roti Safe. V1-V5, PC and NC denote the cell lines (§ Figure 3.12). Digits indicate the clone numbers. Control reactions were performed with the vector and genomic DNA of the empty GS115 strain (N1, N2). The clones without a signal in PCR2 comprised single insertion clones, which could be used for further analysis.

Three validated single insertion clones were retained for each variant to serve as biological replicates in the following experiments. The clones are denoted in Table 3.1.

Table 3.1. Retained single insertion clones of the *GAP* RMCE MCL variants for the expression analysis.

Name of transformed vector	Clone numbers
V1. GS115/pYTA-RudolphRFP	1, 2, 3
V2. GS115/pYTAΔF3-RudolphRFP	1, 2, 3
V3. GS115/pYTAF3PR-RudolphRFP	1, 2, 3
V4. GS115/pYTAatgF3-RudolphRFP	2, 3, 4
V5. GS115/pYTAUR-RudolphRFP	1, 2, 3
PC. GS115/pGAPZAΔ8-RudolphRFP	1, 2, 3
NC. GS115/pYTA (empty)	1, 2, 4

3.1.4.4 Expression analysis of the modified RMCE master cell lines to evaluate the influence of FRT F_3 on gene expression

The negative influence of FRT F_3 on gene expression in *P. pastoris* should be evaluated on two comparatively evaluated layers. First, the production of RudolphRFP was quantified using flow cytometry. Secondly, the transcription levels of the *RudolphRFP* gene were assessed by quantitative reverse transcription PCR (RT-PCR). The analyses were performed with three biological single insertion clones of the modified RMCE MCLs V1 to V5 as well as the control cell lines (§ 3.1.4.3, Table 3.1). Each biological clone was used in technical quadruplicates. Hence, 84 samples were analyzed in total (four replicates x three clones x seven variants).

3.1.4.4.1 High-throughput expression analysis of RudolphRFP

For the expression analysis, all clones were cultivated in 96-deepwell-plates with liquid buffered minimal dextrose (BMD) medium. The red fluorescence was measured after 48 h of expression by flow cytometry (§ 2.7.3). The averaged red fluorescence (x-mean) over the counted cell population was used as the measure to compare the samples. The data analysis was conducted using the mean value of the respective technical quadruplicates and the biological triplicates ($RF_{mean;Rud}$). Student's t-tests were performed to validate the statistical significance of the data. A graphical depiction of the results is shown in Figure 3.14.

To compare the relative red fluorescence of the samples, the $RF_{mean;Rud}$ of the PC was set to 100% by definition and the remaining values were normalized accordingly. The auto-fluorescence of *P. pastoris* leaks into the red channel of the flow cytometer. Hence, the $RF_{mean;Rud}$ of the NC was set to 0 by definition (i. e. subtracted from each of the other values), as this variant does not harbor the *RudolphRFP* gene. As expected from the previous experiments (§ 3.1.4.1), no expression was measurable from V1. In contrast, the expression was fully restored in V2 and V5 compared to the PC. Like the PC, V2 and V5 do not harbor FRT F_3 between *GAP* promoter and *RudolphRFP*. The variants V3 and V4 only show a partially restored expression. To fully interpret these results, it was necessary to assess the transcription levels of *RudolphRFP* through RT-PCR first.

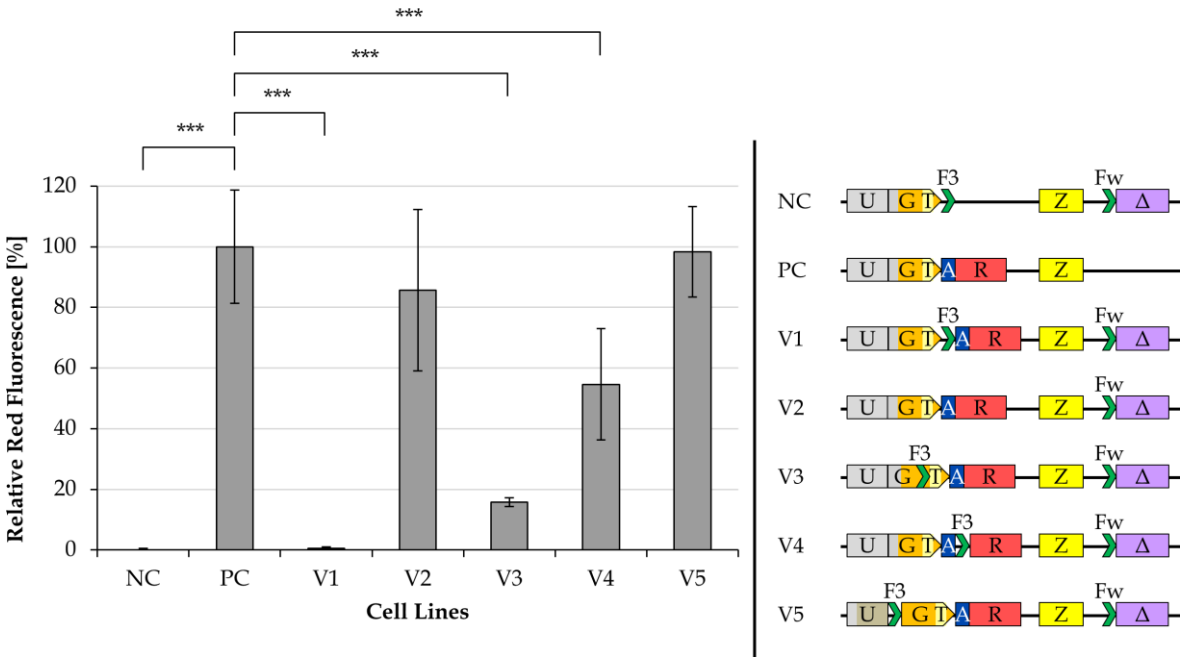


Figure 3.14. Relative red fluorescence of the different RMCE MCL variants.
Mean red fluorescence was measured by flow cytometry after 48 h of expression using four technical replicates of three biological clones of each variant (V1 to V5) as well as, including positive control (PC) and negative control (NC). The fluorescence values per clone were averaged and normalized to the PC and the NC (defined as 100% and 0%, respectively). The results are shown on the left-hand side. RudolphRFP was not expressed in V1. V2 and V5 show comparable expression to the PC. V3 and V4 display reduced expression. A scheme of the cassettes is shown on the right-hand side. U: URG, G: *GAP* promoter, T: TATA box, F3/Fw: FRT sites F_3 and F_{WT}, A: ATG start codon, R: RudolphRFP, Z: Zeocin resistance cassette, Δ: Δ*his4* gene. Student's t-test was performed to evaluate statistical significance of the results (* = P < 0.05; ** = P < 0.01; *** = P < 0.001).

3.1.4.4.2 RudolphRFP Transcription Analysis by RT-PCR

The RT-PCR and the data evaluation were performed as described in detail in chapter 2.6.15. Briefly, the same set of 84 clones was grown overnight and the total RNA was isolated. The cDNA was synthesized using an Oligo(dT)$_{18}$ primer to preferentially amplify the poly-adenine tailed mRNA. For every cDNA synthesis reaction, a control was set up without reverse transcriptase (no reverse transcriptase, NRT control). The RT-PCR was set up with the fluorescent marker EvaGreen (Bio-Rad). EvaGreen solely causes a fluorescent signal by interacting with dsDNA. Since the synthesized cDNA is single-stranded, EvaGreen will only generate a signal by binding to newly amplified dsDNA PCR products during the RT-PCR.

The RT-PCR was performed with three primer pairs (listed in 2.6.15.2) to respectively amplify RT-fragments of *RudolphRFP* and the two housekeeping genes as controls: the genes encoding beta-actin (*ACT1*) and phosphoglycerate kinase 1 (*PGK1*). The synthesized cDNAs and the respective NRT controls were individually used as PCR templates. Additionally, control reactions without template (NTC; no template controls) were performed. The fluorescence signal was measured every PCR cycle. Both the NRT and the NTC control reactions did not display significant fluorescence over the course of 40 PCR cycles. By this, two possible contaminations were excluded, which could lead to false positive signals. First, as no signal was obtained for the

NRT control reactions, no genomic DNA was present in the synthesized cDNA samples. Hence, the measured signals were neither caused by EvaGreen interacting with genomic dsDNA nor by genomic DNA serving as a false-positive PCR template. Secondly, the negative NTC reactions proved that the PCR reaction mixes were not contaminated with DNA or oligonucleotides. These contaminations could otherwise also have resulted in a false signal. Therefore, the obtained fluorescence signals of the conducted RT-PCR experiments could directly be correlated to the cDNA templates and by that, to the relative mRNA transcript levels.

The data was evaluated according to the comparative C_T (cycle threshold) method, which gives access to the fold change of the transcription rates of analyzed genes (§ 2.6.15.3). The C_T describes the PCR cycle in which the fluorescent signal of a sample surpasses the threshold level that is defined by the background fluorescence. Thus, the C_T marks the cycle at which the PCR products become detectable. The more of the amplifiable PCR template for a primer set is originally present in the cDNA, the earlier the C_T is reached. The composition of the cDNA reflects that of the (m)RNA it was synthesized from. Hence, the C_T provides a measure for the amount of transcript in a sample. For this reason, the C_T value allows putting different samples into relation to each other.

The detailed calculations for the data analysis are presented in chapter 2.6.15.3. For the calculations, the C_T of the NC was defined as 40, because it did expectedly not show any signal for the mRNA of *RudolphRFP* (not present in this cell line). In the first step, the readouts of *RudolphRFP* were normalized per sample. Thereby, the readouts for the mRNA of the control gene *PGK1* served as a reference, resulting in the ΔC_T value. The ΔC_T values of the *RudolphRFP* transcripts in each tested cell line (V1 to V5 and the NC) were then put into relation to the ΔC_T of the PC, which was set to 1 (i. e. 100 %) by definition. This resulted in the respective $\Delta\Delta C_T$ values, which were used to calculate the relative fold change of the *RudolphRFP* mRNA in each sample compared to the PC ($2^{-\Delta\Delta CT}$).

Likewise, the fold change of the *ACT1* was calculated to evaluate the uniformity and comparability of the different samples, since *ACT1* (like *PGK1*) should be transcribed at similar levels in all tested cell lines. The results of the transcription analysis are depicted in Figure 3.15. The samples of V1, V2 and V5 displayed similar amounts of *RudolphRFP* transcript compared to the PC. Variants V3 and V4 only showed reduced transcription levels. In contrast to the *RudolphRFP* transcripts, the *ACT1* control showed no significant differences between the cell lines according to Student's t-test. Hence, the obtained results are comparable.

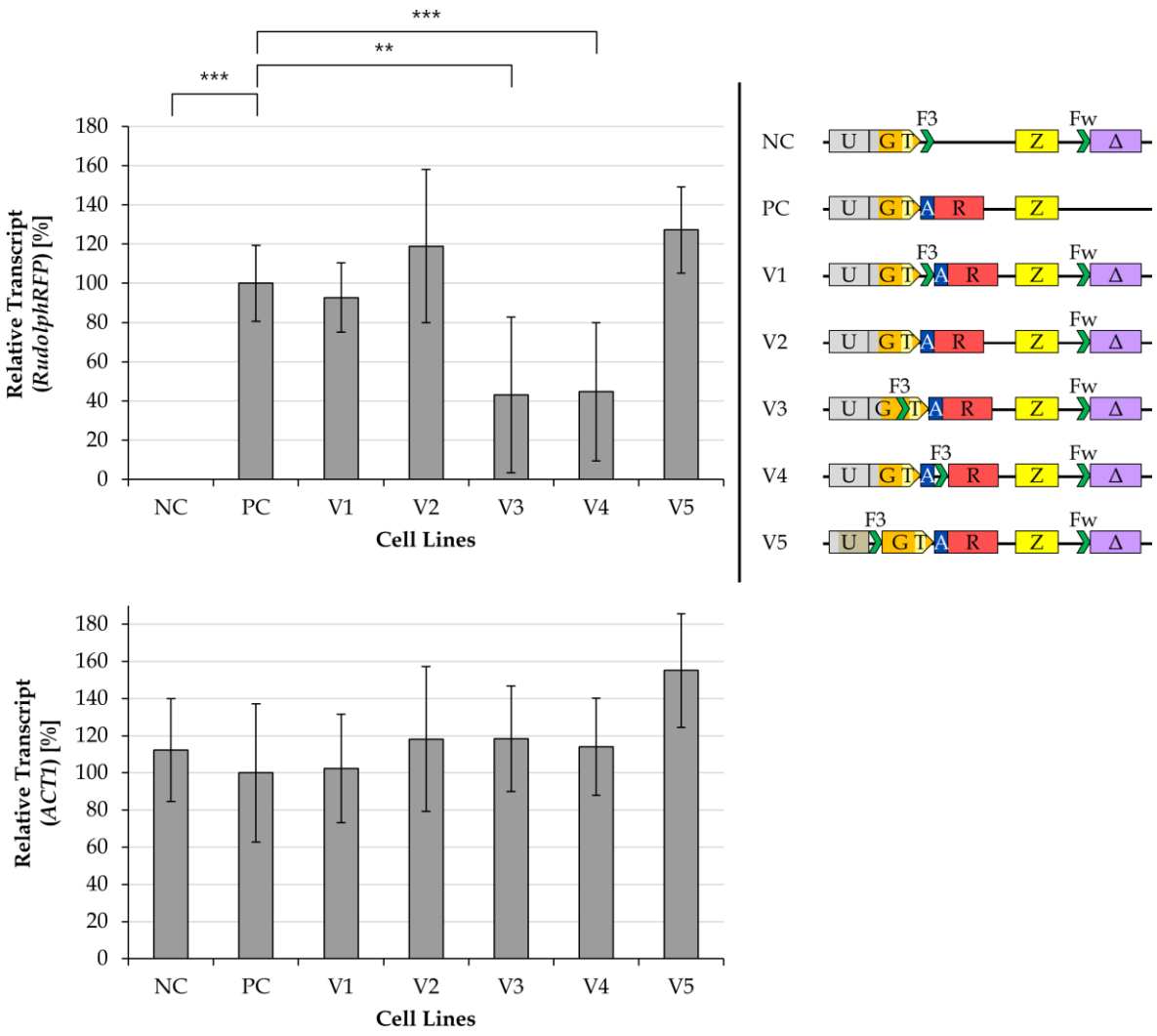


Figure 3.15. Analysis of the *RudolphRFP* mRNA levels in the RMCE MCL variants by RT-PCR.
The RT-PCR was conducted with EvaGreen using the cDNA of each variant (V1-V5) as well as the positive control (PC) and the negative control (NC). The data of a total of 84 clones was analyzed (7 cell lines x 3 biological clones per variant x 4 technical replicates). The obtained values were averaged and the fold change of the transcripts was calculated by the comparative C_T method. Student's t-test was performed to evaluate statistical significance of the results (* = $P < 0.05$; ** = $P < 0.01$; *** = $P < 0.001$). The graph on the top left-hand side displays the fold change ($2^{-\Delta\Delta CT}$) of the transcripts of the GOI *RudolphRFP* in the different cell lines in relation to the PC, which was defined as 100%. V1, V2 and V5 show comparable mRNA transcription levels to the PC. V3 and V4 show reduced transcription levels. Bottom left-hand side: The same test was performed with *ACT1* as a control gene. No significant differences were observed in the different samples, which proved that the obtained RT-PCR data sets of each sample were comparable. The principal design of the cassettes is denoted on the right-hand side. U: URG, G: *GAP* promoter, T: TATA box, F3/Fw: FRT sites F3 and FWT, A: ATG start codon, R: RudolphRFP, Z: Zeocin resistance cassette, Δ: Δ*his4* gene.

3.1.4.4.3 Evaluation of the combined results of the expression screening and the RT-PCR

The influence of FRT F_3 on gene expression was evaluated using the combined results of the conducted analyses of the expression and transcription of the *RudolphRFP* gene (§ 3.1.4.4.1 and 3.1.4.4.2). As expected from previous observations (§ 3.1.2), RudolphRFP was not expressed at detectable levels from the GS115/V1 RMCE MCL, in which FRT F_3 is located between the *GAP* promoter and the GOI (§ Figure 3.14). Interestingly however, *RudolphRFP* was still transcribed at comparable levels to the PC (§ Figure 3.15). Thus, it can be concluded that the loss of expression in V1 is caused at the level of translation. In contrast to V1, the variants V2 and V5 displayed

comparable expression levels of RudolphRFP in relation to the PC (§ Figure 3.14; Figure 3.15). V2 and V5 harbor identical expression cassettes for RudolphRFP compared to the PC, comprising the combination "*GAP* promoter – GOI – *AOX1* transcriptional terminator". The absence of FRT F_3 in the tagging vector V2 (and thus, the generated cell line V2) marks the only difference to V1. This shows that FRT F_3 is indeed the single, critical genetic element responsible for the observed lack of expression of RudolphRFP. Moreover, the restored expression in V5 also indicates that FRT F_3 was not the exclusively critical genetic element for the diminished expression in the AOX1-GAP RMCE MCL (GS115/pYTAaox-F3-GAP-mCherry; § 3.1.4.2). Instead, it is possible that the lack of expression in this cell line was either caused by the *AOX1* promoter or by a combined effect of several genetic elements. However, this was not further analyzed, as V5 already displayed a fully restored expression of RudolphRFP with a virtually identical setup in terms of the RMCE cassette.

The variants V3 and V4 showed measurable expression of RudolphRFP in contrast to V1. Still, the expression levels were significantly reduced compared to the PC, V2 and V5 (see Figure 3.14). Interestingly, these reduced expression levels coincide with likewise reduced transcription levels (see Figure 3.15). Considering the aforementioned conclusions, it was assumed that the expression was restored in V3 by removing FRT F_3 from the 5'-UTR of the mRNA. However, the insertion of FRT F_3 inside the *GAP* promoter probably resulted in reduced transcriptional activity. In V4, the transcribed FRT F_3 site is located in the 5'-region of the ORF. Speculatively, this might negatively affect mRNA stability, resulting in reduced amounts of transcript. This might also apply to the mRNA of the selection trap Δ*his4* (§ 3.1.3). However, it would be irrelevant in the latter case, as the complementing expression of His4 is sufficient for selectivity. In the end, these findings were not subjected to further analysis, as the reduced expression levels render V3 and V4 unfeasible to establish the RMCE system in *P. pastoris*.

In the end, the RMCE MCL variant V5 presents the most viable alternative for the establishment of the RMCE system, as it features a fully restored expression in the presence of an intact RMCE cassette flanked by FRT F_3 and FRT F_{WT}. Consequently, it was used to perform a first cassette exchange via RMCE.

3.1.5 Evaluation of the exchangeability of the RMCE cassette in the modified *P. pastoris* RMCE master cell line V5

Based on the results of the expression analysis, the RMCE MCL variant GS115/V5 was chosen to establish the RMCE method in *P. pastoris*. The required presence of a promoter inside the RMCE cassette to express the GOI renders GS115/V5 incompatible to the originally envisioned RMCE donor vector pFlpBtM. Therefore, a *P. pastoris* exclusive RMCE donor vector had to be cloned. This was done by modifying pYTAUR-RudolphRFP. In the first step, *RudolphRFP* and the Zeocin resistance cassette were replaced with a fragment from pYEXs-yeGFP comprising the *y-eGFP* gene and the "TEF1 promoter – ATG start codon" combination. The vector product was named pYEXsUR-yeGFP, as it harbors the complemented Δ*his4* gene, similar to the previously described pYEXs vectors used for the evaluation of Δ*his4* (§ 3.1.3). In the next step, Δ*his4* was deleted from the vector backbone in order to prevent false-positive clones during the RMCE experiment. The final RMCE donor vector was named pYEXUR-yeGFP (§ Figure 3.16). The detailed cloning steps are described in chapter 2.3. The vector map of pYEXsUR-yeGFP is shown in Appendix IV.5.

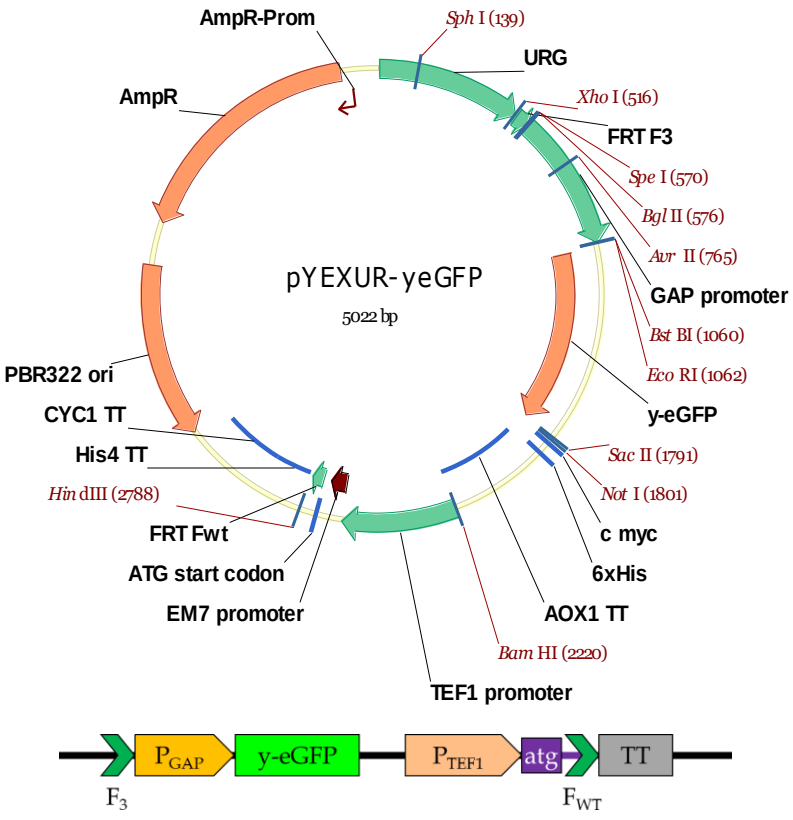


Figure 3.16. Vector map of the RMCE donor vector pYEXUR-yeGFP.
The vector is based on pYTAUR-RudolphRFP. It comprises a pBR322 ori and an ampicillin resistance cassette. The RMCE cassette is flanked by FRT F_3 and F_{WT}. It includes *y-eGFP* as a model GOI and the TEF1/EM7 promoter and an ATG start codon to complement the selection trap upon cassette exchange. The Δ*his4* gene was deleted from the vector. The image at the top was created with Vector NTI Suite 8 (Thermo Fisher Scientific). Teal arrows: eukaryotic promoters and URG. Dark red arrows: prokaryotic promoters. Orange arrows: coding sequences and origin of replication (ori). Blue bars: transcriptional terminators (TT). Below, a depiction of the exchange cassette is shown.

In addition to the donor vector, a preliminary helper vector was cloned to deliver the required Flp recombinase for the RMCE reaction. A detailed description of the vector generation is provided in chapter 2.3. Briefly, the helper vector was designed based on the episomal PARS1 plasmid pBGP1 (§ 2.2). Since pBGP1 is based on pGAPZα-A (Thermo Fisher Scientific), it features the Zeocin resistance gene *Sh ble* for primary selection in *P. pastoris*. The RMCE MCL is already Zeocin resistant from the transformed tagging vector. Therefore, the *Sh ble* gene on pBGP1 was exchanged for the *Tn903* kanamycin resistance gene to enable selection on the antibiotic Geneticin (G418). Furthermore, the MF-α secretion signal was deleted from the vector. The resulting vector was named pBGP1KΔ8. Finally, the *Flp* gene from pOG44 (Thermo Fisher Scientific) was cloned into the vector, resulting in the final helper vector pBGP1KΔ8-FlpWT (Figure 3.17).

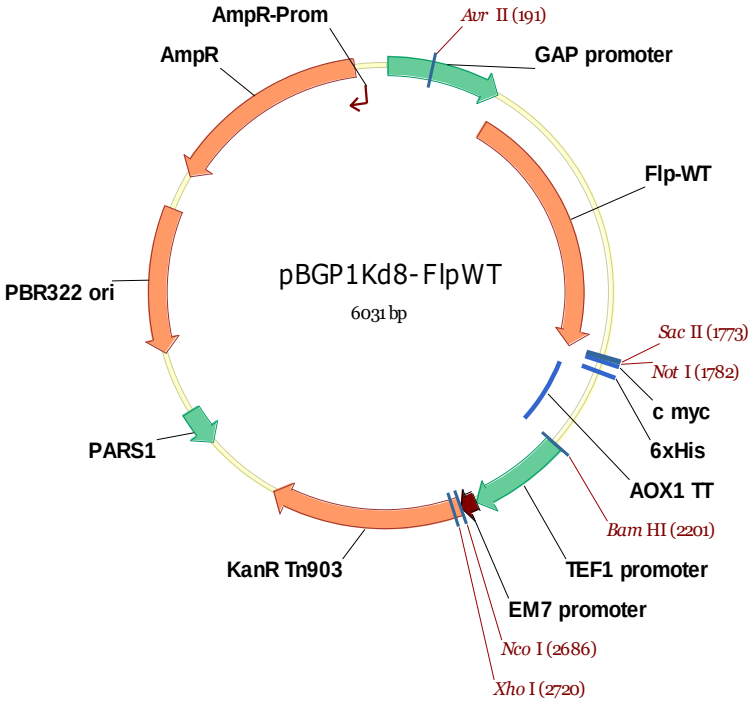


Figure 3.17. Vector map of pBGP1KΔ8-FlpWT.
The vector is based on pBGP1. It comprises a pBR322 ori, an ampicillin resistance cassette and a PARS1 sequence for episomal replication. The Flp gene is expressed constitutively from this plasmid under the control of the *GAP* promoter. In place of the original Zeocin resistance gene, then kanamycin resistance gene Tn903 inserted for direct selection with G418 in *P. pastoris*. The image was created with Vector NTI Suite 8 (Thermo Fisher Scientific). Teal arrows: eukaryotic promoters. Dark red arrows: prokaryotic promoters. Orange arrows: coding sequences and ori. Blue bars: transcriptional terminator sequences. The letter d in the displayed vector name stands for Δ.

To perform the RMCE, both pYEXUR-yeGFP and pBGP1KΔ8-FlpWT were co-transformed into the *P. pastoris* RMCE MCL V5 (§ 3.1.4.4). For the initial experiment a mass ratio of 1:4 (20 ng + 80 ng) was applied. As a negative control for the reaction, each vector was singly transformed into the RMCE MCL. In addition, one transformation was carried out with MQ-H_2O instead of vector DNA. A positive control was generated using the vector pYEXsUR-yeGFP, which harbors the complemented Δ*his4* gene. The vector was linearized and transformed into *P. pastoris* GS115 (§ 2.6.10) for the generation of an exchange simulation cell line by homologous

recombination, as described in chapter 3.1.3. The transformed cells were selected by growth on MD (minimal dextrose) agar plates with 300 µg/mL G418. The cells were also plated in parallel onto MDH agar plates (MD supplemented with histidine) as an additional control.

All transformation samples showed growth on the MDH plates. This proved that all transformation samples after the electroporation still included living cells. For the positive control strain (GS115/pYEXsUR-yeGFP), over 100 colonies were grown on MD. No growth was observed for any of the negative control samples transformed with the individual vectors or MQ-H_2O on MD plates. In contrast, for the cells that had been co-transformed with both vectors (RMCE reaction) four colonies were formed on the MD agar plate. The four colonies were purified and thereafter respectively dropped onto fresh MD and MDH agar plates. The RMCE MCL V5 (GS115/pYTAUR-RudolphRFP) and untransformed GS115 cells were included in this experiment as controls. The agar plates after 48 h of growth are shown in Figure 3.18. Only the four putatively exchanged clones could grow on MD. Furthermore, these four clones showed a green staining, indicating the expression of y-eGFP. In contrast, GS115 and GS115/pYTAUR-RudolphRFP could only grow on MDH. GS115/pYTAUR-RudolphRFP displayed a magenta staining due to the expression of RudolphRFP.

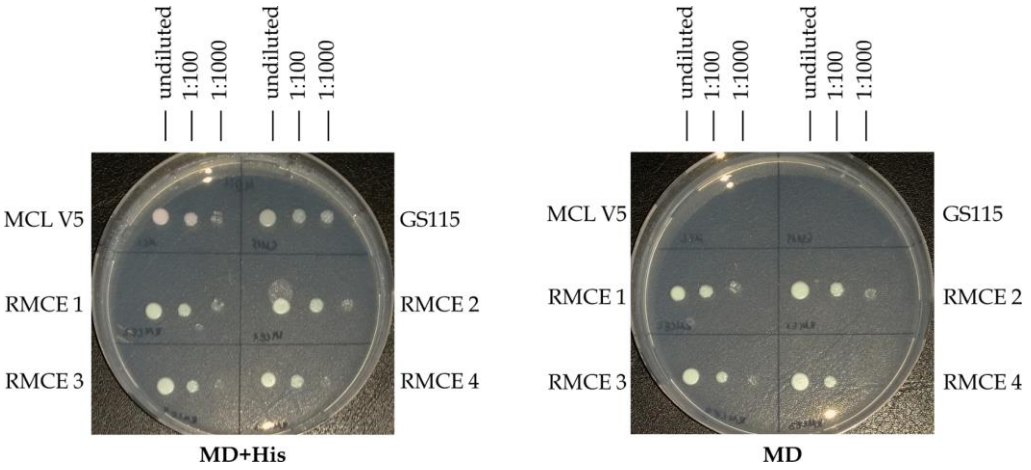


Figure 3.18. Evaluation of the exchanged RMCE producer clones by growth on selective minimal medium. The experiment was performed with the four putatively exchanged producer clones, which were generated through the RMCE reaction (RMCE1-4). In addition, the RMCE MCL V5 (GS115/pYEXsUR-yeGFP) and untransformed GS115 cells were included in the experiment as controls. The cells were respectively dropped onto agar plates with minimal dextrose medium (MD) and MD supplemented with histidine (MD+His) in three different dilutions. Growth was documented after 48 h at 30 °C. Only the four exchanged clones could grow on MD. The RMCE MCL V5 exhibited magenta staining. The clones RMCE1-4 showed green staining.

A preliminary flow cytometric analysis of these clones was conducted (§ 2.7.3). Untransformed *P. pastoris* GS115 cells served as a negative control. The RMCE MCL V5 (GS115/pYTAUR-RudolphRFP) and a visually green fluorescent clone of the positive control transformation agar plate (GS115/pYEXsUR-yeGFP) were also measured for comparison. All four putatively exchanged clones exhibited green fluorescence instead of their original red fluorescence. This shows that the RMCE reaction in *P. pastoris* was successful and that the cassettes were exchanged

(§ Figure 3.19). It is notable that the mean green fluorescence of all four clones was of similar magnitude. This indicates that the RMCE indeed leads to the generation of clones with predictable, uniform expression results in *P. pastoris*, as originally desired (§ Table 3.2).

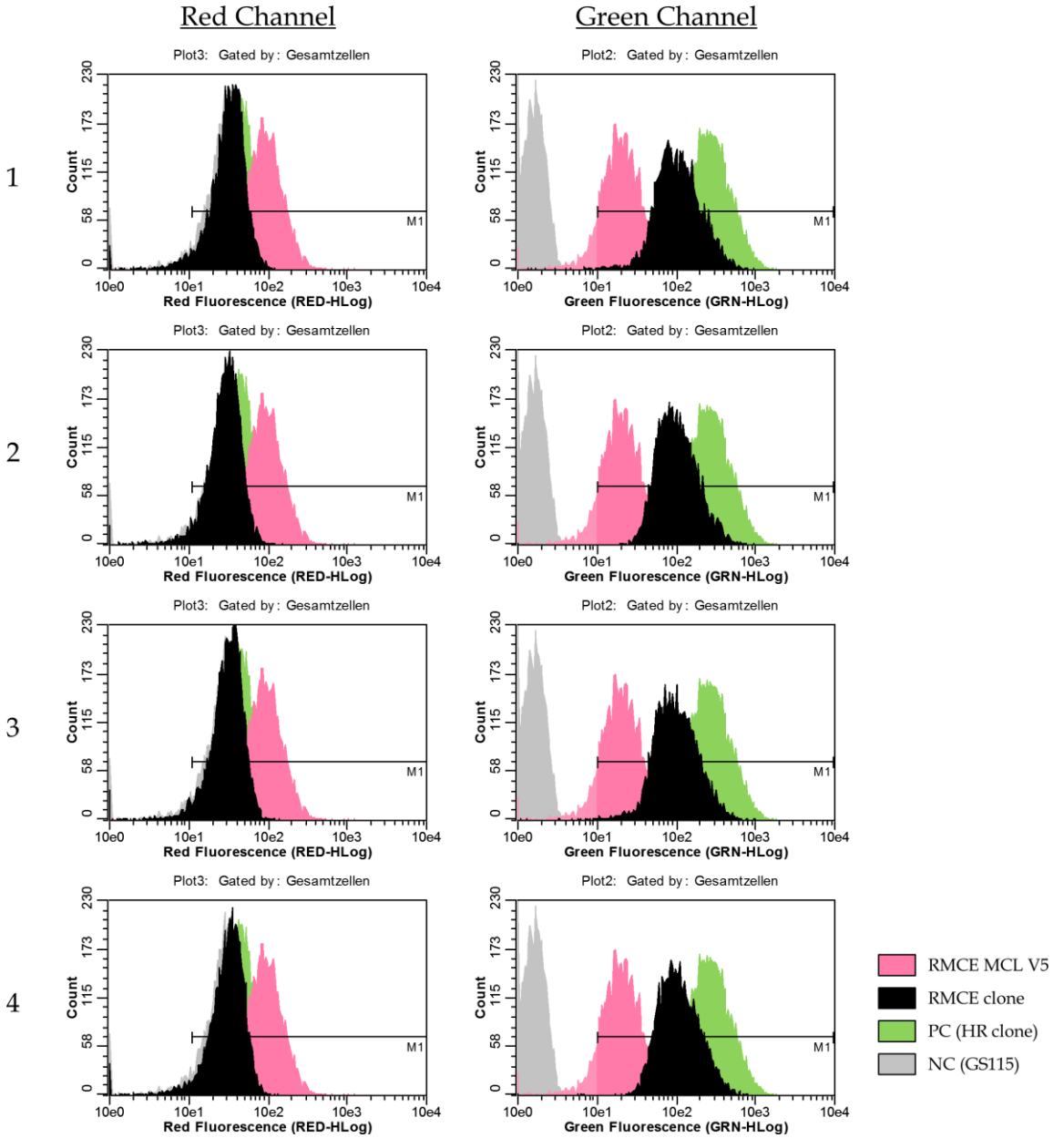


Figure 3.19. Flow cytometry of putatively exchanged RMCE clones.
Red and green channels were detected during the analysis (RED-HLog, GRN-HLog) with a Guava EasyCyte Mini flow cytometer. The events were gated by the living cells (named "Gesamtzellen" in the figure). The counted events are depicted on the y-axis. 10,000 cells were counted per sample. The detected fluorescence on is shown on a logarithmic scale on the x-axis. The RMCE master cell line V5, expressing RudolphRFP, was used as a control (red). Furthermore, "empty" GS115 cells were measured to assess the influence by auto-fluorescence (grey). GS115/pYEXsUR-yeGFP generated by homologous recombination served as a positive control (green). The exchanged clones (black) exhibit green fluorescence, while showing a loss of red fluorescence. The RMCE exchange clones are indicated by the numbers 1 to 4. The fluorescence of RudolphRFP was observed to be leaking into the green channel of the Guava EasyCyte.

The positive control clone (GS115/pYEXsUR-yeGFP) displays a three times higher green fluorescence than the four exchanged clones. As described above, this clone was randomly picked based on its visibly green fluorescence. Assumingly, it presents a jackpot clone. The higher expression level of y-eGFP could be related to multi copy insertions of the vector or to other undefined effects. For example, the random insertion of the vector into the genome could cause unpredictable expression results and might also lead to the increased expression level. However, this was not investigated any further in this PhD thesis.

Table 3.2. Gated mean fluorescence values of the putatively exchanged RMCE clones.

Name	Mean Red Fluorescence	Mean Green Fluorescence
GS115 "empty" (control)	30.93	5.59
GS115/V5 (MCL)	100.5	24.63
GS115/pYEXsUR-yeGFP (PC, HR)	49.35	331.82
GS115/yeGFP-Exchange1	35.11	125.36
GS115/yeGFP-Exchange2	35.13	126.24
GS115/yeGFP-Exchange3	34.16	121.29
GS115/yeGFP-Exchange4	34.59	124.25

All in all, the RMCE reaction in this experiment (albeit not yet optimized) provides a proof of principle for the functionality of the RMCE system in *P. pastoris*. The system could be established using a modified RMCE cassette (V5), which harbors the promoter to drive the expression of the GOI downstream of FRT F_3. The presented model system, which was designed and established in this work, provides a basis for the future generation of *P. pastoris* RMCE MCLs. These new RMCE MCLs could for example be generated by random tagging in order to introduce the RMCE tagging cassette into favorable, yet uncharacterized genomic loci for stable protein production in *P. pastoris*. Notably, the model GOI *y-eGFP* is uniformly expressed in the four exchanged producer clones, which indicates that the productivity is predictable. Thus, the RMCE system would circumvent the need for extensive expression screens, which is up to date a time-consuming drawback of recombinant protein expression in *P. pastoris*.

3.2 Secretory expression of mouse Tmprss2

P. pastoris is a powerful expression host, particularly for the secretory expression of proteins like proteases and for the production of membrane proteins. The previous section of this work (§ 3.1) dealt with the optimization of the expression host *P. pastoris* for the efficient generation of producer cell lines. This section deals with the production of a difficult to express target protein - mouse Tmprss2. Tmprss2 is a type II transmembrane protease (TTSP) present in the airway epithelium. Tmprss2 was shown to be an essential key player for influenza A virus (IAV) infection. The viral surface protein hemagglutinin (HA) mediates the entry of the IAV into the host cells only after proteolytic activation. Tmprss2 is the critical protease *in vivo* to activate monobasic HAs such as H1 (§ 1.7.1). Consequently, Tmprss2 deficient mice were protected from IAV infection. At the same time, these mice did not show any discernible phenotype compared to the wildtype. This makes Tmprss2 an interesting drug target for structure-based drug design (SBDD).

The aim of this work was to establish a reliable protocol to produce sufficiently high amounts of mouse Tmprss2 for crystallization and structural analysis. The whole ectodomain of Tmprss2, which contains the catalytic serine protease domain, was selected for secretory expression (§ 1.7.2). For simplification, the extracellular domain / ectodomain of mouse Tmprss2 will be referred to as "Tmprss2" from here on, unless specifically stated otherwise. To perform SBDD, the structural resolution of the interaction surface between Tmprss2 and its viral substrate HA H1 is of great interest. To enable co-crystallization of Tmprss2 and HA H1, the protein was additionally expressed in a mutant form with diminished catalytic activity (D343N). The expression of the active "wildtype" form of Tmprss2 (Tmprss2-WT) and Tmprss2-D343N was tested in several eukaryotic expression hosts to find the optimal system for the production.

3.2.1 Initial expression of Tmprss2 in *Pichia pastoris*

The expression of Tmprss2 in *P. pastoris* was conducted under the control of the strong, methanol-induced *AOX1* promoter in combination with the mut^S strain KM71H. The recombinant Tmprss2 construct comprises the ectodomain of the native, full length protein that ranges from amino acid 109 to 490 (gene position 325 bp to 1470 bp). The mutant form contains a single point mutation at position 1027 bp of the full length gene, which translates to amino acid 343. The wildtype gene harbors the codon GAC, which codes for aspartic acid (Asp, D). In D343N, this codon is changed to AAC, which codes for arginine (Arg, N). The plasmid pPICZα-A (Thermo Fisher Scientific) was selected as an expression vector for *P. pastoris*. In addition to the *AOX1* promoter, pPICZα-A contains the MF-α secretion signal sequence (MF-α ss) to mediate the secretion of the target protein. For detection and purification, a C-terminal 1x Flag and 8xHis double tag was fused to

Tmprss2. Upstream of the tags a TEV protease cleavage site was inserted to allow for the removal of the tag prior to crystallization. A scheme is of the construct depicted in Figure 3.20.

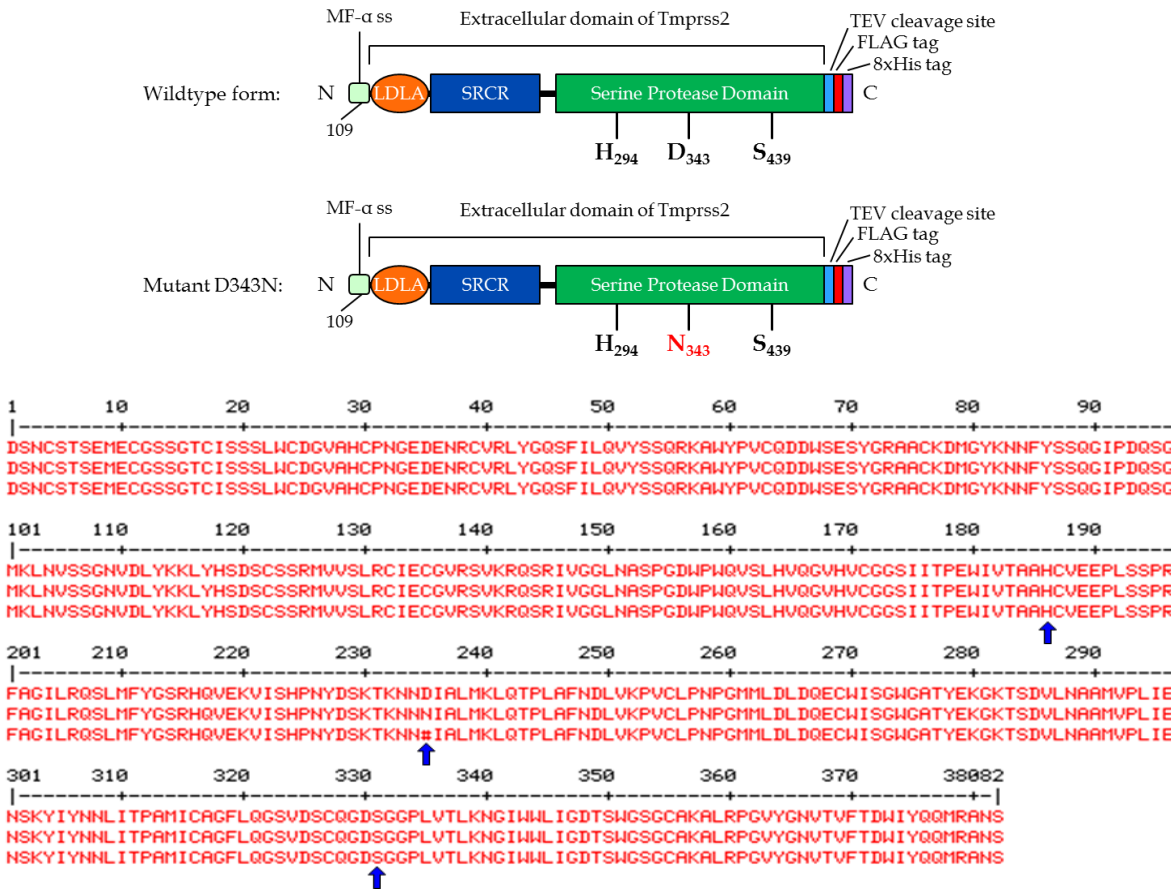


Figure 3.20. Schematic overview of the Tmprss2 constructs for expression in *P. pastoris*.
Top: The mating factor α secretion signal (MF-α ss) is fused N-terminally to the protein to direct it to the supernatant. The mutant form features an amino acid replacement of the aspartic acid of the catalytic triad to asparagine (D343N). C-terminally, a Flag and an 8xHis tag are fused to the protein for purification and detection. Upstream of the tags, a TEV protease site is located to cleave the tags off. LDLA: Low-density lipoprotein receptor class A, SRCR: Scavenger receptor cysteine-rich domain. Bottom: Sequence alignment of the translated sequences of Tmprss2 used in the recombinant constructs (from position 109 of the full protein). The amino acids belonging to the catalytic triad are marked by blue arrows. The mutation D343N is visible at position 235 of the alignment figure.

The *Tmprss2* constructs were cloned into the MCS of pPICZα-A, resulting in the vectors pPICZα-Tmprss2-WT and pPICZα-Tmprss2-D343N. The detailed cloning strategy is described in chapter 2.3. The expression vectors were linearized and respectively transformed into *P. pastoris* KM71H. Primary selection was conducted on YPDS Zeocin (100 µg/mL) agar plates. The single clone isolation was performed with an initial number of 2 x 24 clones, which were spread on multiple YPD agar plates with different concentrations of Zeocin (100, 250, 500, 1000 µg/mL) to select for multi copy clones. The clone numbers were set to use even hundreds (0-99, 200-299, 400-499 etc…) for Tmprss2-WT and odd hundreds (100-199, 300-399 etc…) for Tmprss2-D343N by definition. The selected clones are listed in Table 3.3, ranking by the highest Zeocin concentration still allowing growth of the transformants.

Table 3.3. Survival of the initially analyzed Tmprss2 clones on different Zeocin concentrations.
All transformants showed growth at a Zeocin concentration of 100 µg/mL.

Highest concentration of Zeocin survived [µg/mL]	Tmprss2-WT Clones	Tmprss2-D343N Clones
1000	17	116, 118
500	18, 19, 20	117, 119
250	21	102-115, 121-124

3.2.1.1 Small-scale test expression of Tmprss2 in *P. pastoris*

The WT clones 16 to 21 and the D343N clones 115 to 120 were selected for an initial expression test, as these two sets included clones that were able to survive on 1000 µg/mL (17, 116 and 118), 500 µg/mL (18, 19, 20, 115, 117 and 119), 250 µg/mL (21, 115) and 100 µg/mL (16, 120) Zeocin.

The expression of Tmprss2 was tested in a 50 mL scale (250 mL preculture), as described in chapter 2.8.1. The expression was conducted for 72 hpi (hours post induction). To maintain the induction, methanol was added every 24 hpi to a final concentration of 0.5 % (v/v). The harvested, clear supernatant was dialysed against 50 mM sodium phosphate buffer, pH 7.4 with 300 mM NaCl (§ 2.8.7). Subsequently, Tmprss2 was purified by affinity capture with MagneHis magnetic beads (Promega) and eluted into 50 µL of elution buffer (§ 2.8.9). The samples were analyzed by SDS-PAGE and Western blots (§ 2.9). Figure 3.21 depicts the results of the SDS-PAGE and the Western blot of the elution fractions. It was observed that *P. pastoris* secretes soluble Tmprss2 to the supernatant. Furthermore, the protein is likely glycosylated, judging from the smear of the protein bands towards higher molecular weights, which was particularly visible for the ~45 kDa band of Tmprss2-D343N.

The affinity capture of Tmprss2 from 50 mL of culture supernatant into an eluate of 50 µL equals a 1000-fold concentration. Even then, Tmprss2-WT is only visible on the Western blot. This indicates an overall very low expression level of the protein, which is also considerably lower than the expression level of Tmprss2-D343N. Among the tested clones, the ones that had survived on higher Zeocin concentrations were also producing higher amounts of Tmprss2. Clones 17 and 116 were respectively shown to deliver the strongest expression of Tmprss2-WT and Tmprss2-D343N (Figure 3.21). It is conceivable that the higher copy number of the inserted vector (which goes hand in hand with a higher resistance against Zeocin) resulted in the higher product yields.

The Western blot (see Figure 3.21B) furthermore allowed to draw conclusions about the desired knockdown of the protease activity of Tmprss2 by the mutation D343N. As a putative zymogen, Tmprss2 would cleave itself upstream of the serine protease domain, resulting in two split fragments of ~15 kDa and ~30 kDa. These fragments usually stay linked by a conserved disulfide bond. Under the denaturing conditions of the SDS-PAGE preceding the Western blot, two bands were expected. First, the unprocessed protein should form a band at ~45 kDa (neglecting the

contribution to the mass by glycosylation). Secondly, the processed be visible at ~30 kDa, as only the C-terminal fragment can be detected by the presence of the tag (Figure 3.21A). The observed bands on the Western blot fit to these expectations (Figure 3.21B). Tmprss2-WT showed a dominant band at ~30 kDa (blue arrow in Figure 3.21). This indicates that the protein is virtually completely processed. In contrast, Tmprss2-D343N was mostly unprocessed, as evident by the visibly dominant band at ~45 kDa (black arrow in Figure 3.21) and only a faint signal at ~30 kDa.

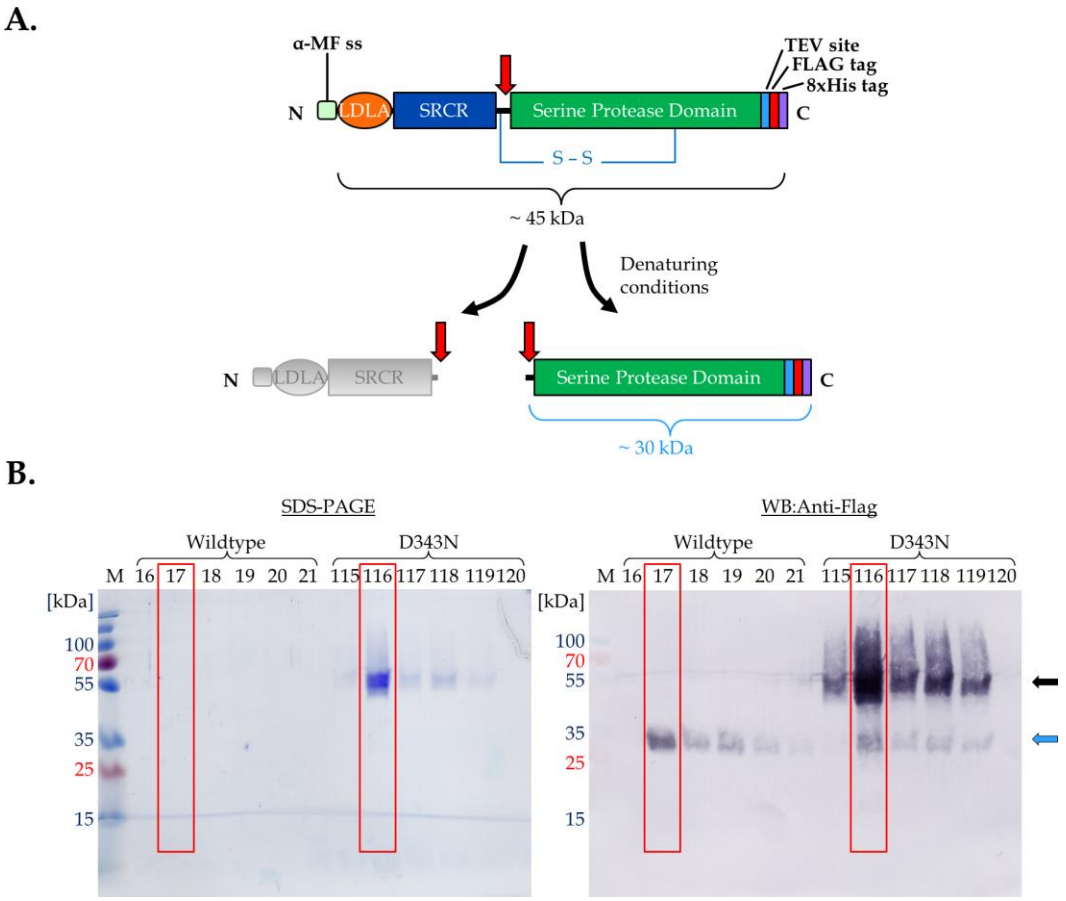


Figure 3.21. Tmprss2 test expression in *P. pastoris*.
(A) Zymogenic cleavage of Tmprss2 should result in two fragments of ~15 kDa and ~30 kDa under denaturing conditions as opposed to the unprocessed form (~45 kDa). On Western blots, only the C-terminal fragment is detectable due to the location of the tags targeted by the primary antibody. **(B)** The wildtype form is completely processed (16 to 21, blue arrow), while the mutant form is mostly unprocessed (115 to 120, black arrow). Clones 17 and 116 showed the strongest expression levels. Tmprss2-WT is expressed at significantly lower levels than Tmprss2-D343N. Even the expression from clone 17 is not detectable on the stained SDS gel. The SDS gel was stained with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric AP staining was performed with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

The fact that the mutant form D343N is mostly unprocessed indicates that the interference or processing activity through host proteases is negligible. On the one hand, these observations confirm that the 30 kDa band of Tmprss2-WT is highly likely the result of the autocatalytic activation of Tmprss2. On the other hand, the mutant form D343N indeed led to a major knockdown of the protease activity of Tmprss2.

The optimal time to harvest the culture post methanol induction was evaluated. For this, the clones 17 (WT) and 116 (D343N) were respectively cultivated in 100 mL expression cultures (from 500 mL of preculture) as described in chapter 2.8.1. Samples were taken every 24 h for a total time of six days post induction. Total cell extracts were generated by mechanical cell disruption (§ 2.8.4). The culture supernatant as well as the soluble and insoluble cell lysate fractions upon centrifugation were analyzed for the presence of Tmprss2 by Western blot (see Figure 3.22).

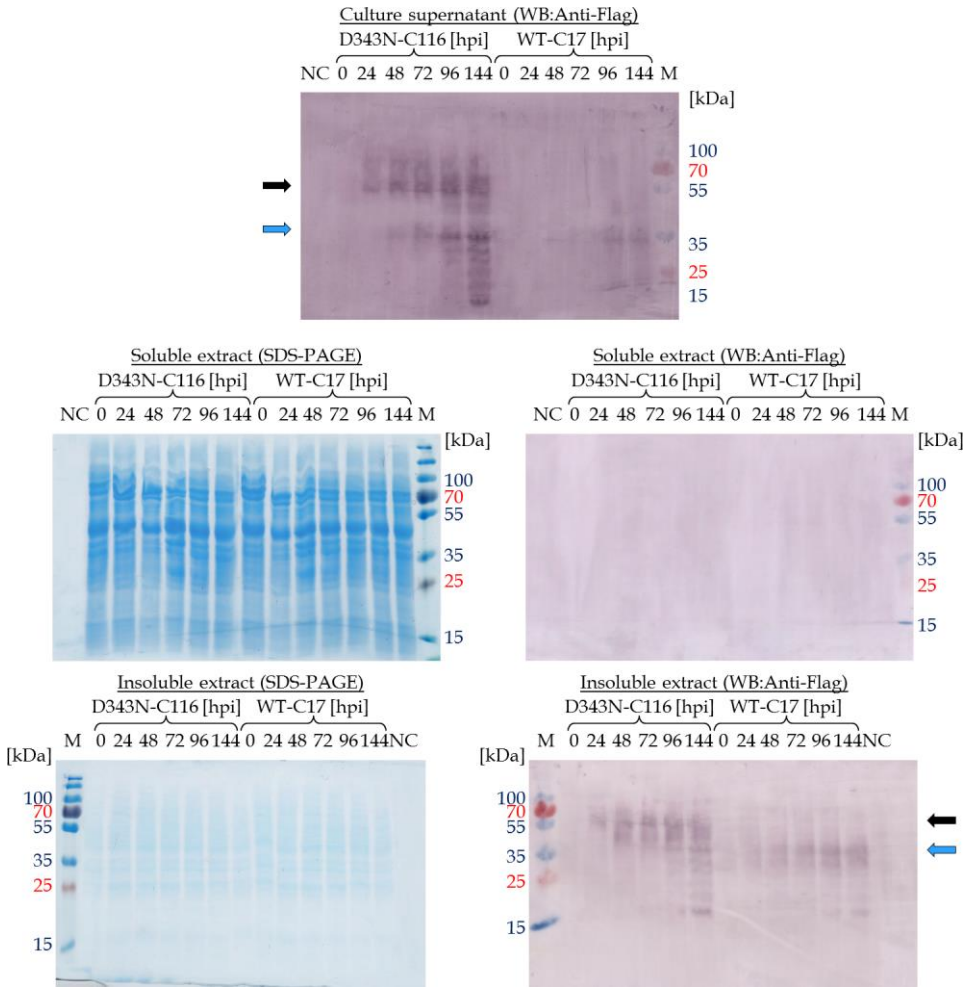


Figure 3.22: Expression test to evaluate the amount of Tmprss2 secretion and optimal harvesting time.
The expression was conducted in 100 mL shake flask cultures for 144 hpi (hours post induction) with clones KM71H/pPICZα-Tmprss2-WT 17 and KM71H/pPICZα-Tmprss2-D343N 116. SDS-PAGE and Western blots were performed of all fractions (culture supernatant, soluble and insoluble cell extract fractions). Upper line: Western blot, supernatant. Second line: SDS-PAGE and Western blot, soluble extract. Lower line: SDS-PAGE and Western blot, insoluble extract. The SDS gels show uniform signals and thereby, comparability of the samples. Tmprss2 was detected in the Western blots. Tmprss2-D343N was showed two bands at ~45 kDa (black arrow) and ~30 kDa (blue arrow) in the supernatant. Tmprss2-WT only showed a band at ~30 kDa. Starting from 72 hpi first signs of degradation of the target protein in the supernatant are visible. For the cell extracts, Tmprss2 was only detected in the insoluble fraction. Tmprss2-WT was expressed at significantly lower rates than Tmprss2-D343N. Staining of the SDS gels was performed with InstantBlue. The Western blot was conducted with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric AP staining was performed with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

Starting from 72 hpi, the supernatant samples show visible degradation of Tmprss2. For this reason, 48 hpi was selected as the optimal time to harvest the supernatant. The cell extract samples showed only faint signs of the target protein in the insoluble fraction, which was expected due to the protein being transported through the secretory pathway of *P. pastoris*. It was concluded that the low yields of Tmprss2 are not caused by significant amounts of improperly folded Tmprss2 getting stuck inside the ER of the cells.

3.2.1.2 Screening for "jackpot" clones (high producers) of Tmprss2 in *P. pastoris*

Following the initial test expression experiments, it was assessed, whether higher expressing clones for Tmprss2 could be identified. To do this, a high-throughput expression screen of 70 additional clones each for Tmprss2-WT (25 to 94) and Tmprss2-D343N (125 to 194) was conducted. An overview of these clones is presented in Table 3.4.

Table 3.4. Survival of the second batch of analyzed Tmprss2 clones on different Zeocin concentrations.
All transformants showed growth at a Zeocin concentration of 100 µg/mL.

Highest concentration of Zeocin survived [µg/mL]	Tmprss2-WT Clones	Tmprss2-D343N Clones
1000	35, 71, 87, 93	170, 183, 187, 194
500	62, 89, 91, 94	126, 130, 177
250	34-39, 63-66,	

The strongest clones of the initial test expression (KM71H/pPICZα-Tmprss2-WT 17 and KM71H/pPICZα-Tmprss2-D343N 116) were included in the experiment as a reference to correlate the expression of the new clones. The expression was conducted in 500 µl BMMY expression cultures for 48 hpi (§ 2.8.1). For a fast, relative evaluation of the expression of Tmprss2, slot blots were performed (§ 2.9.4). Briefly, the culture supernatants were respectively harvested after the expression and directly blotted onto a PVDF membrane. Thereafter, immunostaining was conducted to detect Tmprss2, as described before for the Western blots (§ 3.2.1). The results are depicted in Figure 3.23.

None of the new clones produced higher amounts of Tmprss2 than clones 17 (Tmprss2-WT) and 116 (Tmprss2-D343N), which were consequently designated as the maximal producer clones. As observed before, the expression tends to scale roughly with the vector copy number, evident by the ability of the clones to survive on higher Zeocin concentrations. Congruently with the results from the initial test expression, Tmprss2-WT was produced at significantly lower amounts than Tmprss2-D343N in all screened clones. The only difference between the two proteins is the single point mutation that causes the diminished protease activity of Tmprss2-D343N. Thus, the different expression levels of the two constructs could be explained by the fact that Tmprss2-WT is still active. Tmprss2-WT might be interacting in some way with the host cell environment, thereby negatively influencing its own production or its own half-life.

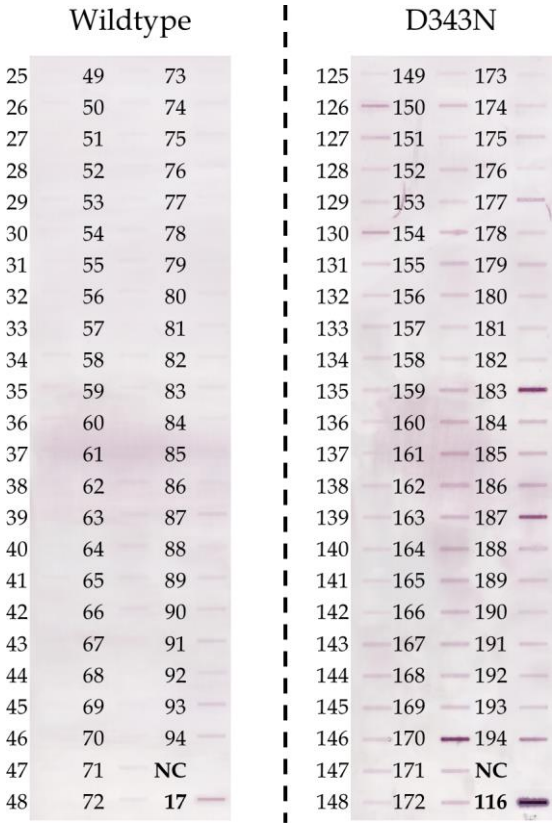


Figure 3.23: Expression screening of Tmprss2 in *P. pastoris* using slot blot.
Clones of KM71H/pPICZα-Tmprss2-WT (25-94) and KM71H/pPICZα-Tmprss2-D343N (125-194) were screened in high-throughput for the expression of the respective Tmprss2 form. The expression was conducted in 500 µL BMMY for 48 hpi. 300 µL of the culture supernatant were loaded onto the slot blots. Clones 116 (D343N) and 17 (WT) were included in the expression screen as comparative controls. The negative control/NC (supernatant from untransformed cells) did not show any signal. This indicates no cross interference by natively secreted host proteins with the immunostaining. None of the screened clones was stronger than clones 116 and 17 for the expression of Tmprss2. The expression patterns correlate to multi insertions of the vector. The slot blot was conducted with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric AP staining was performed with NBT and BCIP.

3.2.2 Analysis of the quantity and quality of Tmprss2 expressed in *P. pastoris*

3.2.2.1 Large-scale production of Tmprss2 in *P. pastoris*

The previous experiments showed that Tmprss2 is difficult to express in high amounts in *P. pastoris*. Since high amounts of pure protein are needed for crystallization, the yields of Tmprss2 from the current *P. pastoris* jackpot clones were quantified to obtain a basis for further improvements. For this purpose, the expression was scaled-up to 2x 1 L shake flask scale cultures (10 L of pre-culture) in BMMY medium with clones 17 (WT) and 116 (D343N), as described in chapter 2.8.1. The concentrated cultures had comparable cell densities at 0 hpi (OD_{595} of 50 and 56 for the cultures of clones 17 and 116, respectively). The expression was conducted for 48 hpi. The cell-free supernatant was concentrated to 500 mL and diafiltered for Ni-NTA affinity capture of Tmprss2 (§ 2.8.7). Tmprss2-WT and Tmprss2-D343N were respectively captured using a 1 mL Ni-NTA column (§ 2.8.11). The eluent fraction was collected in a series of 1 mL fractions. SDS-PAGE and Western blots of the fraction samples were performed. Tmprss2 was successfully captured as shown in Figure 3.24. In contrast to the mutant form, the purified Tmprss2-WT was only detectable on Western blot. Tmprss2-D343N elution fractions 3 to 8 were pooled and dialyzed to remove imidazole. For Tmprss2-WT, the elution fraction 4 was separately dialyzed, while the

remaining elution fractions from 2 to 10 were additionally pooled for dialysis. After the dialysis, the protein concentration was measured with a Nanodrop (§ 2.6.14).

Tmprss2-D343N was quantified at 0.3 mg total (0.15 mg/L). Tmprss2-WT was not quantifiable from elution fraction 4. Even by pooling all Tmprss2-WT elution fractions and concentrating them with a Vivaspin 6 (MWCO 5 kDa) column to 50 µL, quantification was not possible.

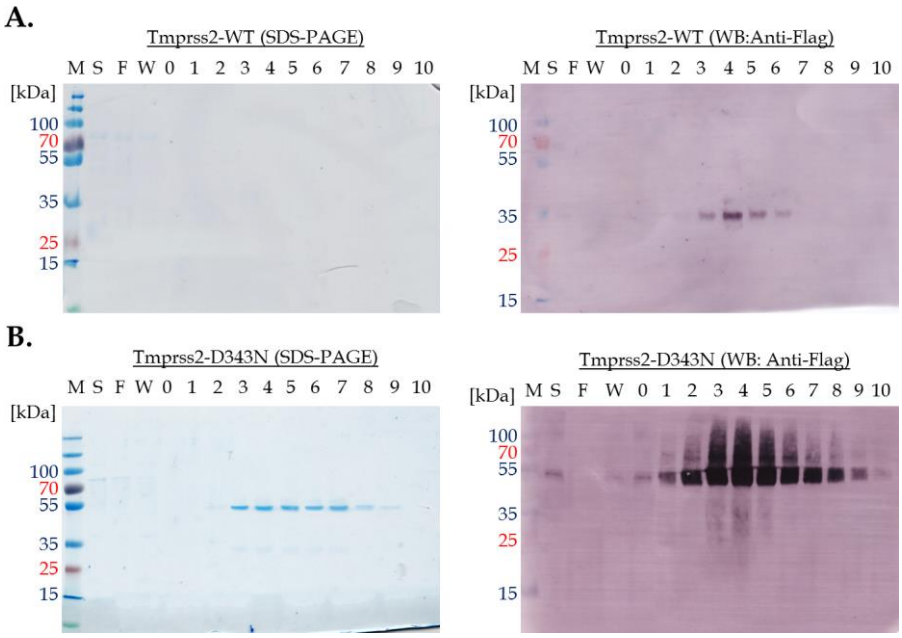


Figure 3.24. Affinity capture of Tmprss2 from 2 L of *P. pastoris* expression culture.
Tmprss2-WT and Tmprss2-D343N were expressed in shake flasks (2 L scale) and respectively captured on a 1 mL Ni-NTA column. SDS-PAGE and Western blot are shown with the harvested culture supernatant (S), the flow through of the Ni-NTA column (F), the wash fraction (W) and the elution fractions (0-10). Tmprss2 was completely captured from the supernatant. **(A)** Tmprss2-WT. The protein was only visible on the Western blot. All elution fractions were retained for further analysis. **(B)** Tmprss2-D343N. The elution fractions 3 to 8 were retained for further analysis. Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric AP staining was performed with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

The cultivation of *P. pastoris* in a bioreactor offers optimized process conditions. Apart from the control of the pH and the temperature, the bioreactor allows for better aeration to cultivate at higher cell densities compared to shake flasks (200 - 300 OD_{595} units/mL). Moreover, by application of a methanol sensor driven methanol feeding pump a defined level of methanol can continuously be maintained in the bioreactor. This allows for a constant and more efficient induction by methanol. It was tested whether these conditions would lead to a quantifiable improvement of the expression yields of Tmprss2 in comparison to shake flask cultures.

The expression was conducted in a 2 L scale in a Labfors bioreactor (§ 2.8.1). The clone KM71H/pPICZα-Tmprss2-D343N 116 was selected to perform the initial experiment, as its expression yields could be directly compared to the shake flask experiments. The methanol induction was maintained for 48 hpi. During the process, the reactor reached an OD_{595} of ~200

units. The total culture supernatant (~1.6 L) was harvested, concentrated to 500 mL and diafiltered for affinity capture (§ 2.8.7). Tmprss2-D343N was captured on a 5 mL HisTALON column using an ÄKTAFPLC (§ 2.8.11). The eluent fraction was collected in a series of 1 mL fractions. The purification process was analyzed by SDS-PAGE and Western blot. The results are depicted in Figure 3.25. Tmprss2-D343N was successfully captured. The elution fractions C1 to C9 were pooled and dialyzed to remove the residual imidazole (§ 2.8.7). Finally, the protein concentration was quantified via Nanodrop (§ 2.6.14) at 0.6 mg total (0.3 mg/L culture volume). This presents a two-fold increased yield compared to the 2 L shake flask culture. The purified material was used to test, whether Tmprss2 could be deglycosylated. Furthermore, it was tested, whether the protein was still intact under native conditions, as this could not be concluded under the denaturing conditions of SDS-PAGE and Western blot.

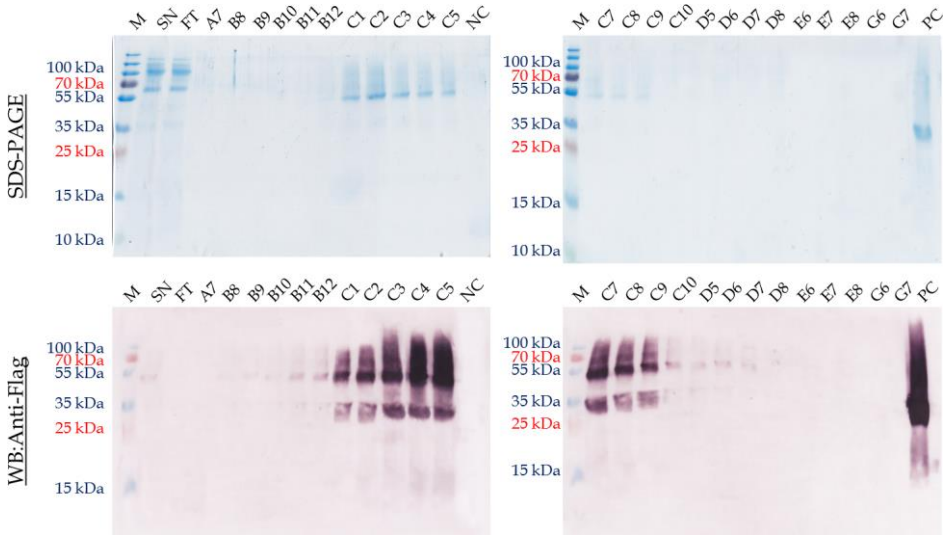


Figure 3.25. Protein purification of Tmprss2-D343N from clone 116 in 2 L bioreactor scale.
The purification of Tmprss2-D343N was analyzed with an ÄKTAFPLC. The samples of different fractions were loaded onto the gel. SN: Concentrated and diafiltered reactor supernatant. FT: Flow through of the HisTALON column. A7: Wash fraction. B8 - G7: Elution fractions, which had shown a UV signal during the ÄKTA run. NC: Negative control (culture supernatant of *P. pastoris* KM71H/pPICZα-A). PC: Positive control; Tmprss2-D343N from *P. pastoris* KM71H/pPICZα-Tmprss2-D343N C116 (processed and partially degraded, but still sufficient as a Western blot control). Staining of the SDS gel was performed with InstantBlue. Western blot detection was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric AP staining was performed with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus. Fractions C1 - C9 contained the target protein and were pooled for further analysis.

3.2.2.2 Deglycosylation of Tmprss2 expressed in *P. pastoris*

A homogeneous glycosylation pattern consisting of small type glycans on the surface of a target protein is generally less flexible, which is beneficial for the formation of protein crystals. In the previously described experiments, a glycosylation smear, assumedly caused by a heterogenous glycosylation pattern, was particularly visible at the ~45 kDa band of the stronger expressed Tmprss2-D343N (§ 3.2.1, Figure 3.21). Therefore, a deglycosylation test was conducted. *P. pastoris* applies glycosylation of the high mannose type. Hence, it was evaluated, whether

glycan cleavage would be possible with Endo Hf (New England Biolabs). For this, 1 µg Tmprss2-D343N was incubated with varying amounts of Endo Hf (§ 2.8.12). SDS-PAGE and Western blot were performed for analysis. The results are shown in Figure 3.26. Endo Hf was able to deglycosylate Tmprss2-D343N under both native and denaturing conditions.

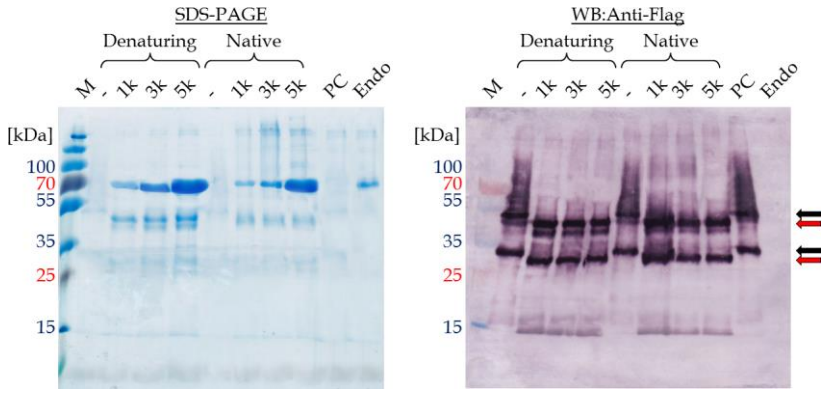


Figure 3.26: Deglycosylation test of Tmprss2-D343N with Endo Hf.
Deglycosylation of ~1 µg Tmprss2-D343N expressed in *P. pastoris* KM71H was conducted according to NEB's manual with 1,000, 3,000 and 5,000 U of Endo Hf (indicated by "1k, 3k, 5k") at 37 °C for 1 h. In addition, a reaction without Endo Hf was set up ("-"). Tmprss2-D343N served as a positive control (PC) that had not been subjected to the reaction mix or to 37 °C. Additionally, a sample only comprising Endo Hf ("Endo") was added as a negative control. Tmprss2-D343N was deglycosylated under both denaturing and native conditions. This is shown by the significantly reduced glycosylation smear and the lower molecular weight (red arrows) in comparison to the glycosylated samples (black arrow). Staining of the SDS gel was performed with InstantBlue. The Western blot was conducted with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

3.2.2.3 Validation of the protein quality of Tmprss2 expressed in *P. pastoris*

The quality of Tmprss2 was validated to ensure that the observed pattern of two bands (~45 kDa and ~30 kDa) on SDS-PAGE and Western blot (§ Figure 3.21) was indeed related to zymogenic self-processing and not to protein degradation. For this, Tmprss2-D343N was analyzed by two methods. First, an intact protein mass determination (ESI-LC/MS) was performed. Purified Tmprss2-D343N was deglycosylated with Endo Hf under native conditions and captured by MagneHis magnetic beads (§ 2.8.9). A quality check of the eluent fraction was performed by SDS-PAGE and Western blot. Thereafter, the buffer of the protein was changed to MQ-H_2O (§ 2.9.5) and the ESI-LC/MS was done by in the MS platform of the research group Cellular Proteomics at the HZI Braunschweig. The results show that Tmprss2 was still intact, despite the - by then - significant ~30 kDa band on the Western blot (Figure 3.27, arrow). Consequently, this indicates that the observed band pattern on SDS-PAGE and Western blot is a result of zymogenic cleavage due to the rest activity of Tmprss2-D343N. Furthermore, it shows that the split halves of Tmprss2 remain covalently disulfide-bonded to each under native conditions.

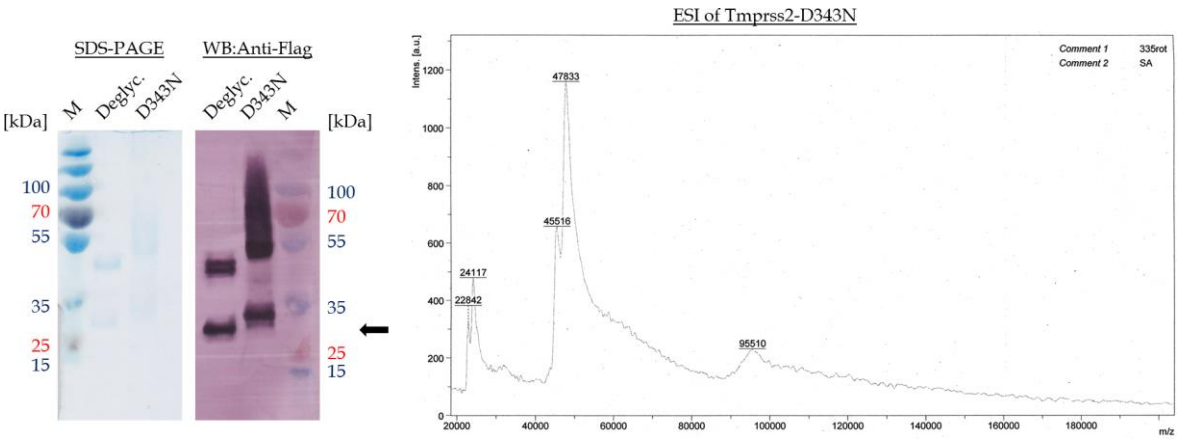


Figure 3.27. Intact ESI-LC/MS of Tmprss2-D343N after deglycosylation with Endo Hf.
Deglycosylation of Tmprss2-D343N was conducted with 5,000 U of Endo Hf for 3 h at 37 °C under native conditions. The deglycosylated sample (Deglyc.) was analyzed by SDS-PAGE and Western blot (left-hand side) in comparison to the glycosylated protein (D343N). Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus. The ESI-MS (right-hand side) showed at strong peak at the size of ~45 kDa. This indicates that Tmprss2-D343N is still an intact protein comprising the full ectodomain.

The integrity of Tmprss2-D343N was additionally tested by Native PAGE. For this, the unquantifiable eluate concentrate of the Tmprss2-WT affinity capture was added to Tmprss2-D343N (~15 µg) at a volumetric ratio of 1:2. A control sample of Tmprss2-D343N was handled equally, but without the addition of Tmprss2-WT. The samples were checked by SDS-PAGE and Western blot. Tmprss2-D343N was completely processed in the presence of Tmprss2-WT, as indicated by the dominant 30 kDa fraction on the Western blot (Figure 3.28, blue arrow). This provides an initial proof of activity, as it is conceivable that Tmprss2-WT has exerted the cleavage of Tmprss2-D343N. It further shows that the rate of the zymogenic activation of Tmprss2-D343N can be significantly increased by the addition of Tmprss2-WT. Native PAGE was performed as described in chapter 2.9.2 with both unprocessed and processed Tmprss2-D343N. Both samples ran at the same size of ~45 kDa (Figure 3.28). This shows that the protein is intact under native conditions, supporting the results of the intact mass determination (Figure 3.27). Hence, the ~30 kDa band under denaturing conditions (SDS-PAGE / Western blot) is indeed caused by the disruption of the internal disulfide bond of processed Tmprss2.

In summary, recombinant Tmprss2 could be successfully expressed in *P. pastoris* KM71H, albeit at low yields. The mutant form D343N was expressed at 0.15 mg/L in shake flasks and 0.3 mg/L in a bioreactor. The active wildtype form was expressed at significantly lower levels and could not be quantified from 2 L of shake flask expression culture. However, it displayed sufficient proteolytic activity to facilitate the zymogenic cleavage of the mutant form. Furthermore, Endo Hf was identified as a feasible endoglycosidase to catalyze the deglycosylation of Tmprss2 expressed in *P. pastoris*. Lastly, the protein quality was verified by intact mass determination and Native PAGE, which showed that Tmprss2 was intact following deglycosylation and proteolytic activation.

The primary aim of this project was to establish a protocol to generate the required amounts of Tmprss2 for further analyses. For this reason, the feasibility of different eukaryotic expression hosts for the production of Tmprss2 was tested, before further attempts were made to improve the expression yield in *P. pastoris*.

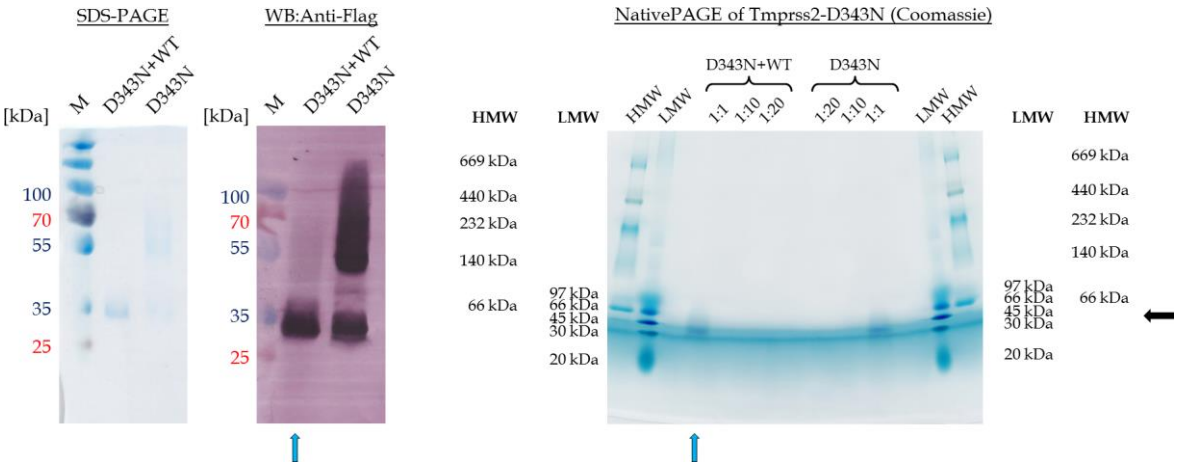


Figure 3.28. Catalytic cleavage and NativePAGE of Tmprss2-D343N.
Left-hand side: Tmprss2-D343N was completely processed through the addition of an unquantifiable amount of Tmprss2-WT after 3 h at 37 °C (D343N+WT). This resulted in a single, dominant ~30 kDa band on SDS-PAGE and Western blot (blue arrow). An equally handled control without the addition of Tmprss2-WT still showed an additional, significant 45 kDa band (D343N). Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus. Right-hand side: The Native PAGE revealed that the processed Tmprss2-D343N is still an intact protein comprising the full ectodomain. The completely processed and the untreated sample respectively form a single band at ~45 kDa (black arrow). In addition to the undiluted samples (1:1), two dilutions were loaded onto the gel (1:10 and 1:20). Only the undiluted sample showed a visible signal. The staining of the Native PAGE was performed with Coomassie G250. Mw standards: Amersham HMW (66 – 669 kDa) and LMW (14.4 – 97 kDa) standards.

3.2.3 Evaluation of eukaryotic cell lines as expression hosts for Tmprss2 using the multi-host expression system

Since both Tmprss2-WT and Tmprss2-D343N were expressed at low levels in *P. pastoris*, the expression was assessed in a number of eukaryotic systems in order to find a better expression host. The expression of Tmprss2 was analyzed in HEK293-6E cells, SF21 and Hi5 insect cells. For the test expression in HEK293-6E cells, the multi-host vector pFlp-BtM-III was used (§ 1.4 and 2.2). It harbors the strong CMV promoter to express the GOI and an EBV oriP for episomal replication of the plasmid and its propagation to the daughter cells during mitosis. For the generation of recombinant EmBacY baculovirus to enable baculoviral expression in insect cells (BEVS), a newly developed derivative of pFlpBtM-III was used (pFlpBtM-III-insect). It harbors the "hr5 – OpiE2 – p10" promoter combination to enhance the baculoviral expression in insect cell lines. The IgG secretion signal sequence, present on both pFlpBtM-III variants, was used to mediate the secretion of Tmprss2 into the culture supernatant.

The expression vectors were generated by respectively cloning the genes *Tmprss2-WT* and *Tmprss2-D343N* (including the tags; § 3.2.1, Figure 3.20) from the generated pPICZα vectors into pFlpBtM-III and pFlpBtM-III-insect (see details in chapter 2.3).

3.2.3.1 Transient expression test of Tmprss2 in HEK293-6E cells

The vectors pFlpBtM-III-Tmprss2-WT and pFlpBtM-III-Tmprss2-D343N were used to transiently transfect 30 mL suspension cultures of HEK293-6E cells (§ 2.7.4). The transfection was monitored by the constitutive expression of GFPq from the co-transformed vector pTTo/GFPq. The expression was conducted for 96 hpt (hours post transfection). Daily samples were taken and analyzed by flow cytometry for GFP fluorescence (§ 2.7.3). About 90 % of the cells expressed GFPq in the experiment, indicating a high efficiency of the transfection. The cells were additionally counted and the vitality was calculated (§ 2.7.1). The cultures maintained high vitality of about 90 % during the experimental run. A graphical overview is depicted in Figure 3.29

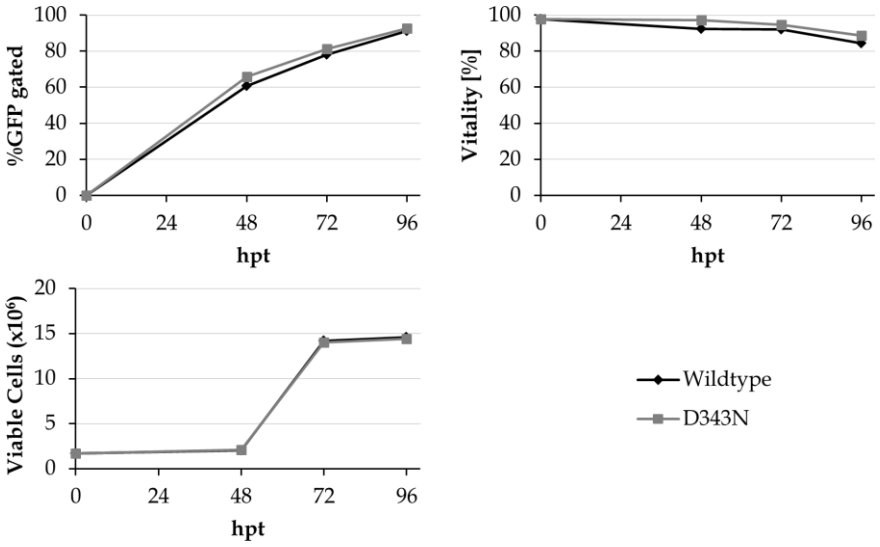


Figure 3.29. Graphical overview of the transfection of HEK293-6E cells for the expression of Tmprss2.
The graphical data for two independently performed experiments is shown. The cells were counted manually with a Neubauer hemocytometer and the vitality was assessed by trypan blue staining. The GFPq expression was monitored by flow cytometry. The data is listed in detail in Appendix III.

The expression was analyzed with 1 mL samples that were taken at 0, 24, 72 and 96 hpt. The total supernatants of the harvested expression cultures were concentrated twenty-fold by TCA precipitation (§ 2.8.6). The cell pellets were lysed according to chapter 2.8.5. The supernatant samples as well as the soluble and insoluble cell lysate fractions were analyzed by SDS-PAGE and Western blot. The results of the experiment are depicted in Figure 3.30. Despite the high level of GFP expression, Tmprss2-WT and Tmprss2-D343N were not expressed at detectable levels. Therefore, the expression of Tmprss2 was not further pursued in HEK293-6E cells, as they did not present a viable alternative to *P. pastoris*.

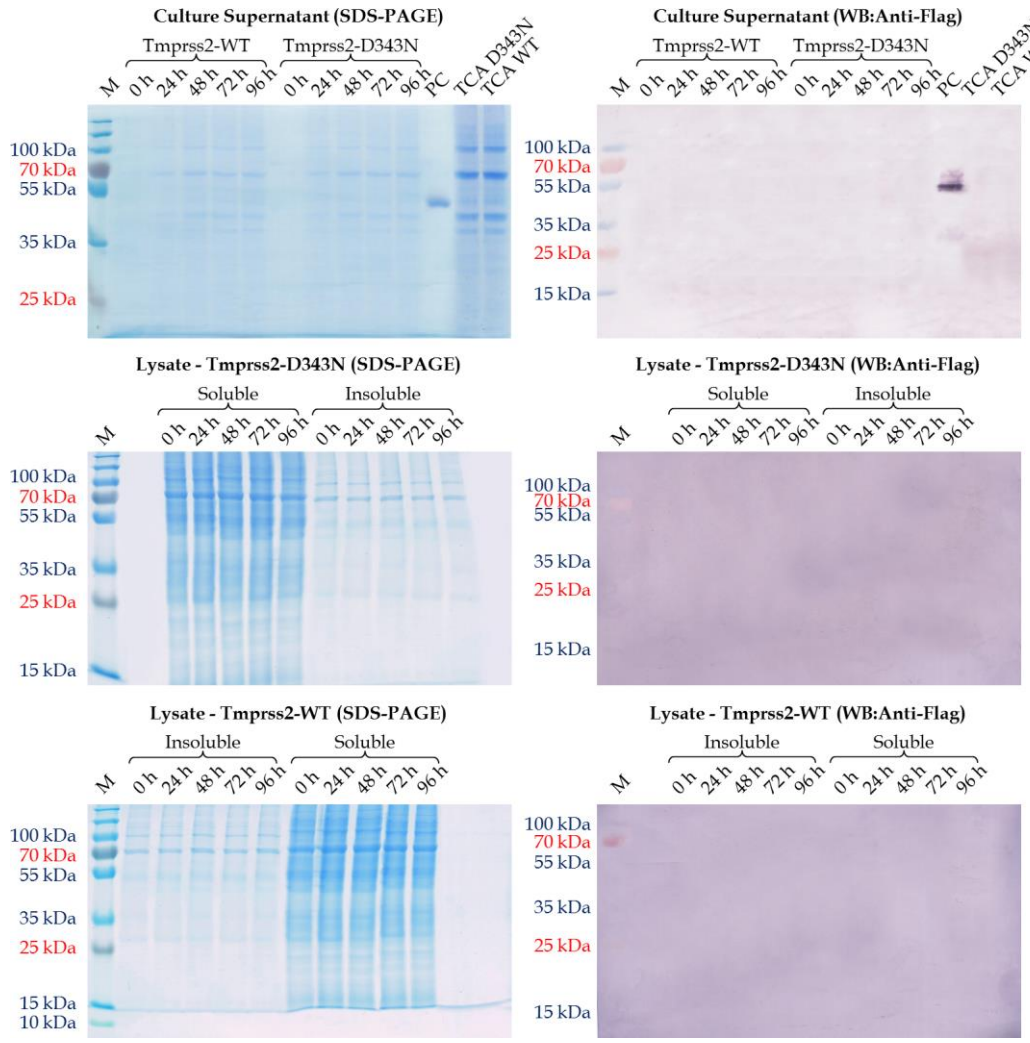


Figure 3.30. Tmprss2 expression analysis of HEK293-6E cells.
Top line: Samples of the culture supernatant (taken at 0, 24, 48, 72 and 96 hpt) were analyzed for the presence of Tmprss2-WT and Tmprss2-D343N by SDS-PAGE and Western blot. Additionally, the harvested total culture supernatant was TCA precipitated and included in the analysis. Moreover, the soluble and insoluble cell lysate fractions of Tmprss2-D343N (middle line) and Tmprss2-WT (bottom line) were analyzed. Neither Tmprss2-D343N nor Tmprss2-WT could be detected in any of the samples. Staining of the SDS gels was performed with InstantBlue. The Western blots conducted with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Tmprss2-D343N from *P. pastoris* KM71H/pPICZα-Tmprss2-D343N clone 116 served as a positive control for the Western blots (PC). Mw standard (M): PageRuler Prestained Plus.

3.2.3.2 BEVS expression test of Tmprss2 in SF21 and Hi5 insect cells

For the test expression of Tmprss2 in insect cells, recombinant bacmids (EmBacY) were generated with the vectors pFlpBtM-III-insect-Tmprss2-WT and pFlpBtM-III-insect-Tmprss2-D343N (§ 2.7.5). The bacmids were transfected into SF21 insect cells to amplify the recombinant baculovirus. The resulting virus stock (virus amplification 1 / VA1) was used at a volumetric ratio of 1:10 to respectively infect 120 mL cell cultures of SF21 and Hi5 (§ 2.8.3). The rate of the infection was monitored by flow cytometry through the constitutive expression of YFP from the bacmid backbone (§ 2.7.3). The cells were counted and the vitality was calculated (§ 2.7.1). Furthermore, a CASY Counter was used to track the increase of the cell diameter, which is a

frequently observed effect of the baculoviral infection (§ 2.7.2). A graphical overview of the data is presented in Figure 3.31.

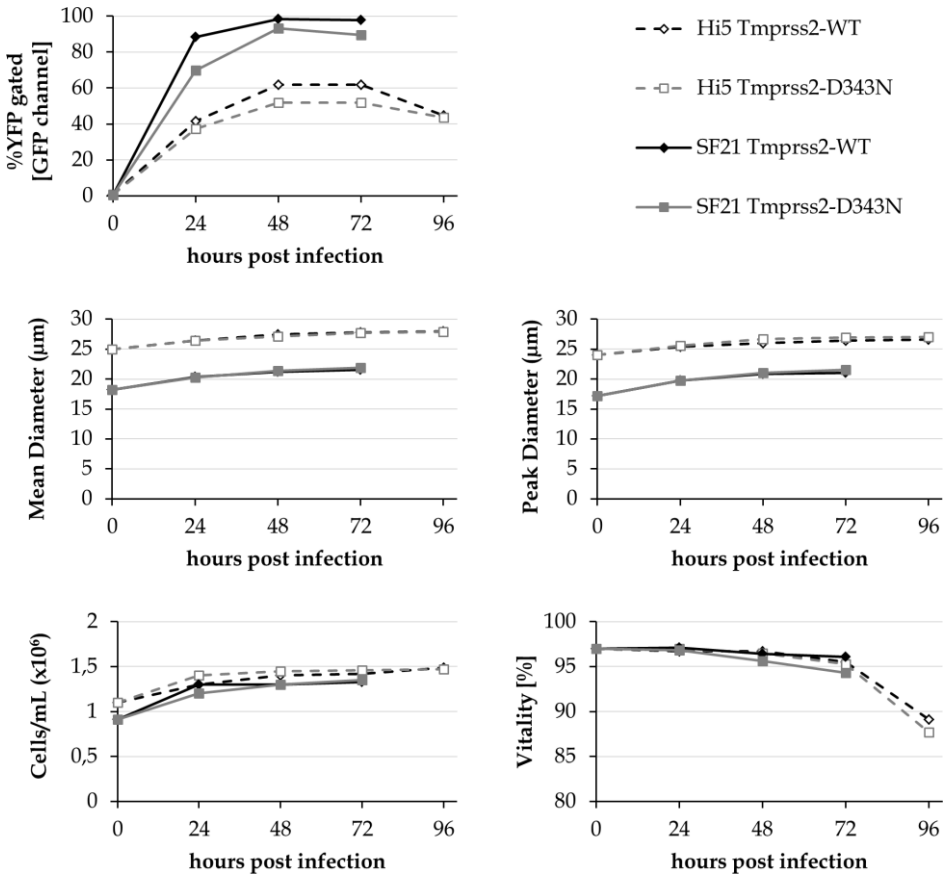


Figure 3.31: Graphical overview of the infection of SF21 and Hi5 cells for the expression of Tmprss2.
The graphical data depicts the daily assessed growth and infection parameters. The cells were counted manually with a Neubauer hemocytometer. The vitality was assessed by trypan blue staining. The YFP expression of EmBacY was monitored by flow cytometry in the GFP channel. The mean and the peak diameter were analyzed with a CASY counter. The infection efficiencies according to the YFP fluorescence were at ~50 % for Hi5 (dotted lines) and >90 % for SF21 (straight lines). After 24 h, an increase in the cell diameter was observed, indicating viral infection. Concomitantly, cell proliferation was halted at 24 – 48 h post infection. SF21 cultures were harvested after 72 hours post infection. Hi5 cultures were cultivated for 24 more hours, until the vitality fell below 90 %. The data is listed in Appendix III.

The respective YFP fluorescence shows that the SF21 and Hi5 cells were infected at efficiencies of at least 50 % (Hi5) and >90 % (SF21), which should be sufficient for the expression of the target protein at detectable levels. Daily samples were taken from the expression cultures for analysis. Both the supernatant and the cells were retained for analysis. Cell extracts were generated of the harvested cells (§ 2.8.5). The expression was terminated at 72 hours post infection (SF21) and 96 hours post infection (Hi5). The total cultures were harvested. To monitor the expression, the supernatants as well as the soluble and insoluble cell extract fractions upon centrifugation were analyzed by SDS-PAGE and Western blot. However, neither the wildtype nor the mutant form of Tmprss2 could be detected, with the exception of a very faint (and hence not reliably strong) signal of ~20 kDa in the insoluble lysate fraction (see Appendix III.3).

Therefore, the Western blot analysis was repeated with chemiluminescent staining to increase the sensitivity (§ 2.9.3). The Western blots were performed with the daily samples of the supernatants and the soluble lysate fractions. In addition, the insoluble lysate fractions of the respective final samples (72 h and 96 h post infection) were included in the analysis. Moreover, 1 mL supernatant samples from the total harvested cultures were concentrated twenty-fold by TCA precipitation (§ 2.8.3) to further enhance the detection. The results of the chemiluminescent Western blots are depicted in Figure 3.32. The Western blots were developed for an extended exposure time of 900 s to ensure the detection of the lowest possible amount of protein. Neither form of Tmprss2 was detectable in the supernatant samples, even after TCA precipitation. The aforementioned ~20 kDa band was clearly detected in the insoluble fraction. Tmprss2 should be detectable at ~45 kDa and ~30 kDa under denaturing conditions (§ 3.2.1). Hence, the ~20 kDa band could present the false-positive detection of a cross-reacting host protein.

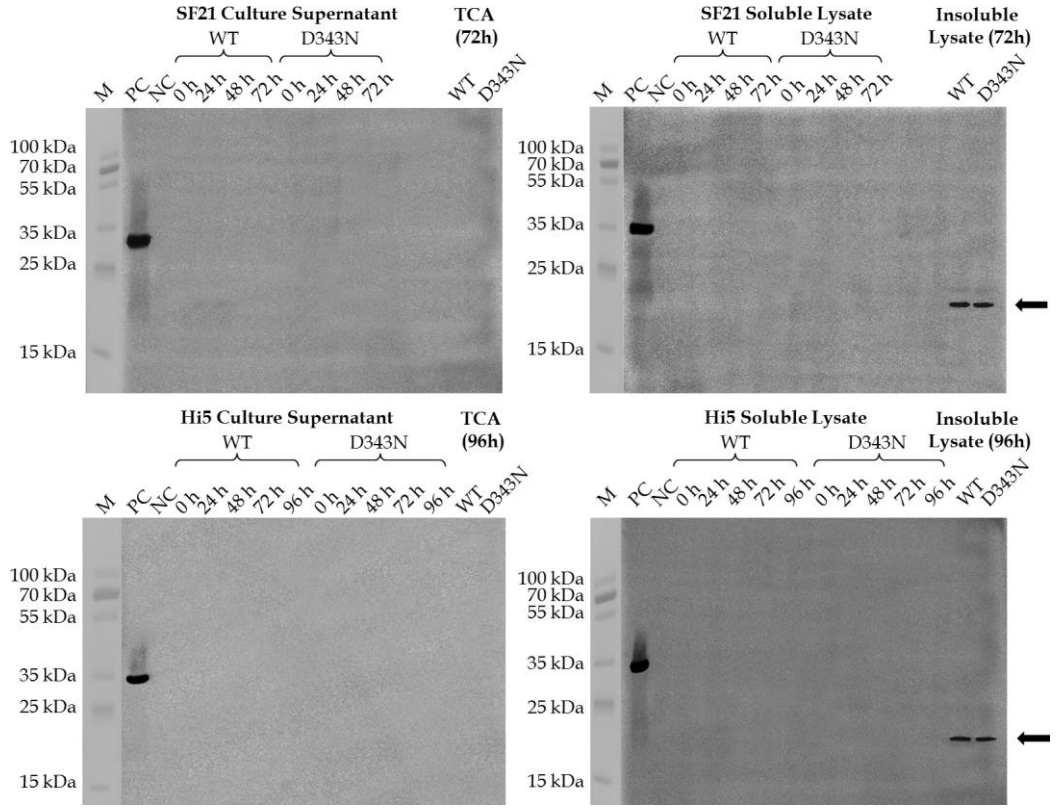


Figure 3.32. Analysis of the Tmprss2 expression in SF21 / Hi5 cells by chemiluminescent Western blots.
The Western blots were developed for an extended time for 900 s. The daily supernatant samples and the respective soluble lysate fractions of the cultures were analyzed. The times (0 – 96 h) indicate the hours post infection. TCA precipitated supernatant from the harvested culture was included in the analysis. The insoluble cell lysate was only loaded for the final sample (72 h or 96 h, respectively). The only observed signal was a 20 kDa band in the insoluble fraction (arrow) that was present for both insect cell lines. The Western blot was performed with a primary mouse α-Flag mAb and a secondary HRP-conjugated α-mouse pAb. Chemiluminescent HRP-staining was done with the Lumi-Light Western Blotting Substrate (Sigma). (Processed) Tmprss2-D343N from *P. pastoris* KM71H/pPICZα-Tmprss2-D343N C116 served as a positive control (PC). Mw standard (M): PageRuler Prestained Plus.

Alternatively, the signal could be related to Tmprss2. This could indicate that Tmprss2 is present in the secretory pathway, as the insoluble fraction includes the ER. However, this was not subjected to further analysis for two reasons. First, the 20 kDa signal was only detected on Western blots, which were extensively developed with a highly sensitive chemiluminescent staining method. This indicates very low levels of the detected protein. So even if the signal would be related to Tmprss2, the expression level in the tested insect cell lines would be too low to generate the required amounts of the target protein. Secondly, in this case, the ~20 kDa band would most likely result from the undesired cleavage of Tmprss2 by host cell proteases.

In summary, the results show that Tmprss2 is not secreted in insect cells at detectable levels, despite sufficiently high efficiencies of the baculoviral infection. Hence, insect cells do not present a viable alternative to *P. pastoris* as an expression host for Tmprss2. Consequently, this approach was not further pursued in favor of improving the expression yields of Tmprss2 in *P. pastoris*.

3.2.4 Optimization of the Tmprss2 constructs for the production in *P. pastoris*

Tmprss2 could not be successfully expressed and secreted in HEK293-6E, SF21 or Hi5 cells (§ 3.2.3). A preliminary test expression of Tmprss2 was also conducted in *E. coli* Rosetta (DE3) with a periplasmic signal sequence. However, Tmprss2 was not expressed at detectable levels (data not presented). In contrast, *P. pastoris* offered proper processing and secretion of Tmprss2 into the supernatant during the previously described experiments, although only low expression yields were obtained (§ 3.2.1 and 2.3). Overall, *P. pastoris* presented the most feasible alternative among the tested expression hosts for the production of Tmprss2. Hence, several optimizations to the target gene constructs were assessed to improve the production yields of Tmprss2 in *P. pastoris*.

3.2.4.1 Expression of the serine protease domain of mouse Tmprss2 in *P. pastoris*

An important aim of this project was to establish a method for the generation of sufficiently high amounts of Tmprss2 for co-crystallization with HA H1. The expression of truncated versions of a target protein can significantly enhance the expression yields, because of the lower complexity of the constructs. It was tested whether the secretory expression of just the serine protease domain (SPD) of Tmprss2 would provide a substantial improvement over the expression of the whole ectodomain (§ 3.2.1).

The SPD of Tmprss2 (amino acids 254 – 490 of the full length protein) was cloned into pPICZα-A (Thermo Fisher Scientific) from the existing expression vectors pPICZα-Tmprss2-WT and pPICZα-Tmprss2-D343N (§ 2.3). Thereby, the C-terminal tags of the recombinant genes were included. The vectors were named pPICZα-Tmprss2-SPD-WT and pPICZα-Tmprss2-SPD-D343N.

The vectors were respectively linearized and transformed into *P. pastoris* KM71H (§ 2.6.10). Primary selection was performed on YPDS agar plates (100 µg/mL Zeocin). Single clones were isolated on YPD agar plates with 100 µg/mL and 1000 µg/mL Zeocin to select for multi copy insertions, which had been shown to be favorable for the expression of Tmprss2 in the previous experiment (§ 3.2.1). Clone numbers were assigned as described previously for the full ectodomain (§ 3.2.1). The prefix "s" was added to distinguish the new clones expressing only the SPD. The expression of 142 clones each was tested in high-throughput (§ 2.8.1). An overview of these clones is listed in Table 3.5. The current jackpot clones 17 (KM71H/pPICZα-Tmprss2-WT) and 116 (KM71H/pPICZα-Tmprss2-D343N) were cultivated in parallel as comparative controls to correlate the expression from the new clones. Furthermore, untransformed KM71H were cultivated in parallel as a negative control. The expression was comparatively tracked by immunostaining of supernatant samples on a slot blot (§ 2.9.4).

Table 3.5. Survival of the analyzed Tmprss2-SPD clones on an increased Zeocin concentration.
All transformants showed growth at a Zeocin concentration of 100 µg/mL.

Highest concentration of Zeocin survived [µg/mL]	Tmprss2-SPD-WT Clones	Tmprss2-SPD-D343N Clones
1000	s6, s28, s34, s79, s88, s94, s201, s235, s239	s106, s124, s127, s136, s137, s138, s159, s195, s195, s301, s316, s322, s323, s335, s336, s339, s341

The results of the slot blot are shown in Figure 3.33. Expressing just the SPD of Tmprss2 did not enhance the yields over the jackpot clones 17 and 116, although the SPD is a significantly smaller and simpler target protein compared to the whole ectodomain. It is possible that the folding of the protein – particularly the SPD – presents the major bottleneck for the expression of Tmprss2 in *P. pastoris*. In this case, the removal of the N-terminal domains would not necessarily result in an increased product yield. Even more so, their removal could also be disadvantageous, if they aid in the folding or stabilization of the SPD. It should be noted that the active form of Tmprss2 was again expressed at significantly lower rates than its mutant counterpart, which was particularly evident for the multi insertion clones. This is congruent with the former observations made for the whole ectodomain (§ 3.2.1). Consequently, it was concluded that the expression of just the SPD did not comprise an improvement over the production of the whole ectodomain of Tmprss2 in *P. pastoris*. Hence, this approach was not further pursued.

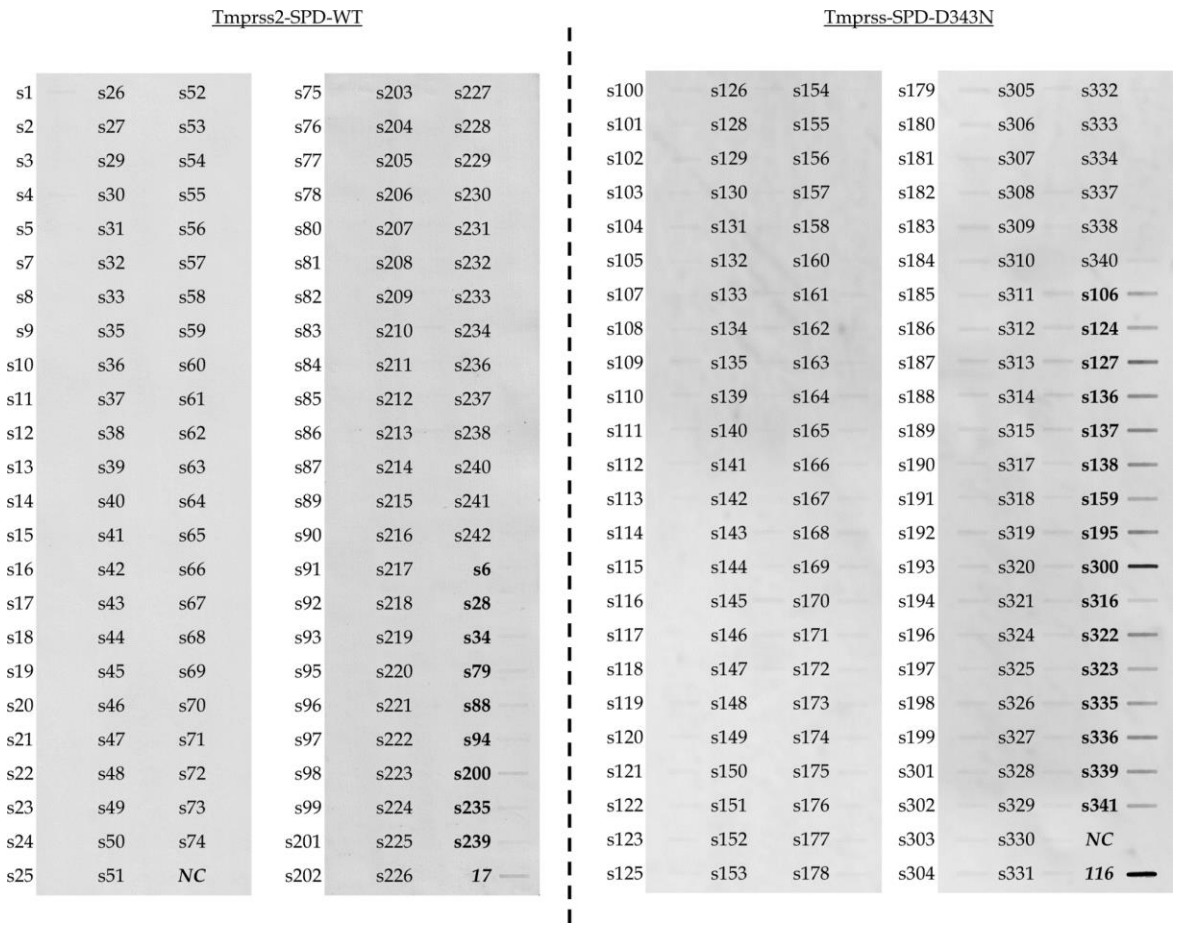


Figure 3.33. Slot blot of the expression screen of Tmprss2-SPD.
Clones of KM71H/pPICZα-Tmprss2-SPD-D343N (s100 – s199; s300 – 341) and KM71H/pPICZα-Tmprss2-SPD-WT (s1 – 99; s200 – s242) were screened in high-throughput for the expression of the respective Tmprss2 form. The expression was conducted in 500 µL BMMY for 48 hpi. 300 µL of the culture supernatant were loaded onto the slot blots. Clones 116 (KM71H/pPICZα-Tmprss2-D343N) and 17 (KM71H/pPICZα-Tmprss2-WT) were included in the expression screen as comparative controls. The negative control/NC (supernatant from untransformed cells) did not show any signal. This indicates no cross interference by natively secreted host proteins with the immunostaining. None of the screened clones was stronger than clones 116 and 17 for the expression of Tmprss2. The mutant form is expressed at higher rates than the wildtype form. The expression levels roughly correlate to multi insertions of the vector, as evident by survival on high concentrations of Zeocin. The slot blot was conducted with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric AP staining was performed with NBT and BCIP.

3.2.4.2 Expression of codon-optimized Tmprss2 in *P. pastoris*

In the next step to improve the expression yields of Tmprss2, gene optimization for *P. pastoris* was performed. A codon optimized *Tmprss2-D343N* gene was ordered from GenScript ("*Tmprss2-D343Nopt*"), including the C-terminal tags used before ("TEV site – 1xFlag tag – 8xHis tag"). The wildtype gene *Tmprss2-WTopt* was generated by fusion PCR to re-introduce the original codon "GAC" (aspartic acid, D) in place of "AAC" (asparagine, N). Subsequently, *Tmprss2-WTopt* and *Tmprss2-D343Nopt* were cloned into pPICZα-A (Thermo Fisher Scientific), as described in detail in chapter 2.3. The vector products were designated as pPICZα-A-Tmprss2-WTopt and pPICZα-A-Tmprss2-D343Nopt.

Each vector was linearized and transformed into *P. pastoris* KM71H (§ 2.8.1). Primary selection was conducted on YPDS agar plates with Zeocin (100 µg/mL). For single clone isolation, the clones were transferred to YPD agar plates with 100 µg/mL and 1000 µg/mL Zeocin. Clone numbers were assigned as described previously for the full ectodomain (§ 3.2.1). The prefix "o" was used to denote the clones harboring the optimized genes. The clones that were able to survive on the higher Zeocin concentration are denoted in Table 3.6.

Table 3.6. Survival of the analyzed Tmprss2opt clones on an increased Zeocin concentration.
All transformants showed growth at a Zeocin concentration of 100 µg/mL.

Highest concentration of Zeocin survived [µg/mL]	Tmprss2-WTopt Clones	Tmprss2-D343Nopt Clones
1000	o5, o8, o13, o16, o17, o20, o41, o50, o72, o84, o201, o202, o224, o225	o103, o104, o107, o108, o111, o112, o115, o116, o119, o123, o139, o155, o159, o167, o187, o199, o304

Initially, 120 clones of KM71H/Tmprss2-pPICZα-Tmprss2-D343Nopt were tested for the expression of Tmprss2-D343N. The respective jackpot clone with the original CDS (KM71H/pPICZα-Tmprss2-D343N 116) was cultivated in parallel for comparison. Untransformed KM71H cells were included in the experiment as a negative control. The expression screen was performed for 48 hpi in high-throughput (§ 2.8.1). The relative protein amounts were analyzed on slot blots (§ 2.9.4). Purified Tmprss2-D343N was used as a positive control for the immunostaining. The results are depicted in Figure 3.34. The new clones harboring the optimized gene were significantly stronger in respect to the Tmprss2 expression levels than the ones with the original CDS of Tmprss2-D343N (§ 3.2.1.2). This is evident by comparing the respective slot blot signals to the control clone 116. Clones o107 and o108 showed particularly strong signals and were retained for further analysis. The same expression screen was also conducted with 140 clones of KM71H/pPICZα-Tmprss2-WTopt. As controls, KM71H/pPICZα-Tmprss2-D343Nopt clone o108, KM71H/pPICZα-Tmprss2-D343N clone 116 and KM71H/pPICZα-Tmprss2-WT clone 17 were cultivated in parallel. Similar to the expression with the original CDS (§ 3.2.1.2), Tmprss2-WT was expressed at considerably lower levels than the mutant form D343N even in the clones harboring the optimized gene. This further supports the hypothesis that the enzymatic activity carried out by Tmprss2-WT may present the bottleneck for its own expression. Still, improved expression levels were observed for multi copy clones, such as clones o5 and o20. Hence, these clones were retained for further analysis.

In the next step, the expression yields of the maximum producer clones were quantified. For this, 1 L BMMY expression cultures (5 L of preculture) were set up with clones o108 (D343N) and o20 (WT) (§ 2.8.1). At 0 hpi, the cultures had an OD_{595} of 53 (D343N) and 58 (WT). The supernatants were harvested at 48 hpi and diafiltered for affinity capture of Tmprss2 (§ 2.8.7). Tmprss2 was captured using a 1 mL Ni-NTA column (§ 2.8.11). The different fractions of the affinity capture

were analyzed by SDS-PAGE and Western blot (Figure 3.35). In contrast to Tmprss2-D343N, the purified Tmprss2-WT was only visible in the Western blot, which is consistent to the lower observed expression levels during the expression screen (Figure 3.34). The elution fractions 1 to 8 for clone o108 and 1 to 5 for clone o20 were pooled and dialyzed to remove imidazole. The codon optimized clone o20 (Tmprss2-WT) had displayed higher productivity compared to the original jackpot clone KM71H/pPICZα-Tmprss2-WT 17 (Figure 3.34). Still, the amounts expressed by clone o20 in the 1 L expression culture could not be reliably quantified using a Nanodrop spectrometer. In contrast, the mutant form D343N expressed from clone o108 could be quantified at ~0.3 mg/L (0.6 mg total). This presents a 100 % increase over the amounts obtained from the original jackpot clone KM71H/pPICZα-Tmprss2-D343N 116 (§ 2.3).

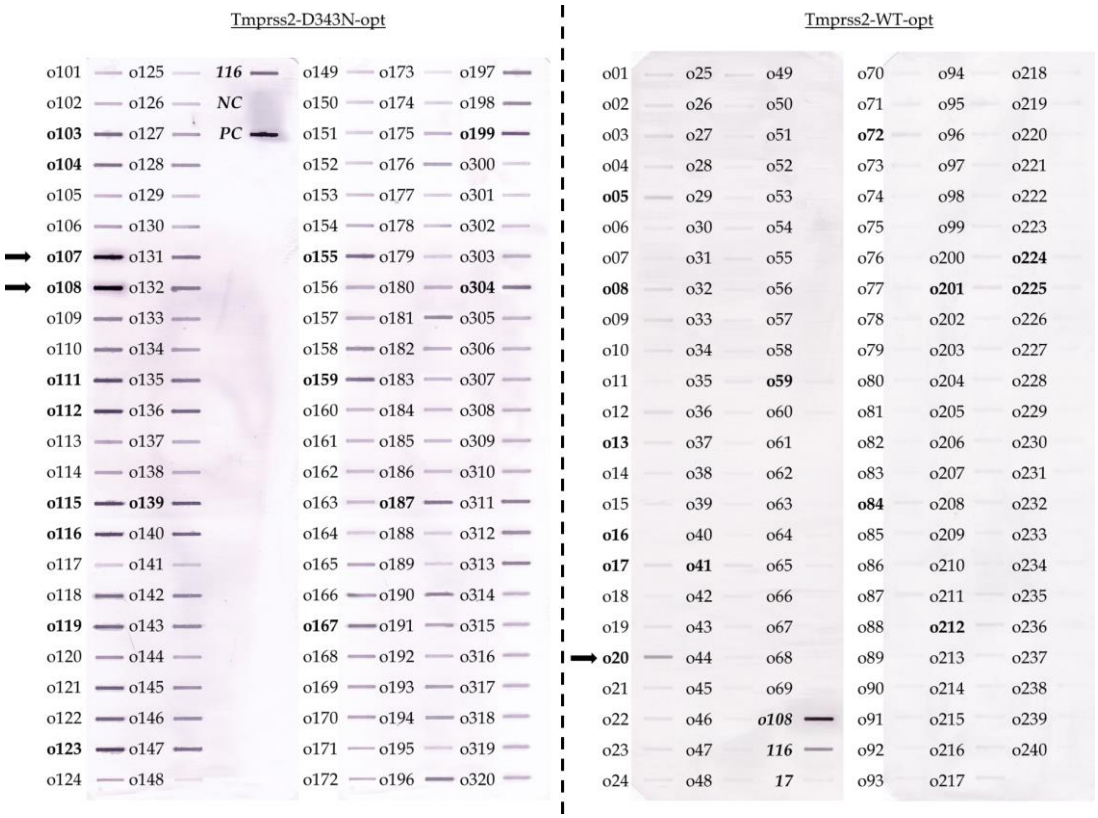


Figure 3.34. Slot blot of the expression screen of Tmprss2opt.
300 µL of culture supernatant from the clones KM71H/pPICZα-Tmprss2-D343Nopt (o100–o199; o300–o320; left-hand side) and KM71H/pPICZα-Tmprss2-WTopt (o1 – o99; o200 – 240; right-hand side) were loaded per slot to assess the expression levels. The optimized clones provide improved expression levels. Clones o107 and o108 were identified as the strongest producing clones for Tmprss2-D343N. Clone o20 was identified as the strongest producing clone for Tmprss2-WT. Generally, the mutant form is expressed at higher rates than the wildtype form. The expression levels roughly correlate to multi insertions of the vector, as evident by survival on high concentrations of Zeocin. Arrows indicate the clones delivering the strongest expression signals. Controls are denoted in bold-italics. They comprised purified Tmprss2-D343N as a positive control (PC) as well as culture supernatant from parallel cultivated clones KM71H/pPICZα-Tmprss2-D343N clone 116; KM71H/pPICZα-Tmprss2-WT clone 17; KM71H/pPICZα-Tmprss2-D343Nopt clone o108 (for the WT screen) and supernatant from untransformed KM71H as a negative control (NC). The blots were stained with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb.

Taken together, the results show that the codon optimization was beneficial for the expression of Tmprss2. However, Tmprss2-WT is still expressed at significantly lower levels than Tmprss2-D343N and could consequently not be quantified from a 1 L expression scale.

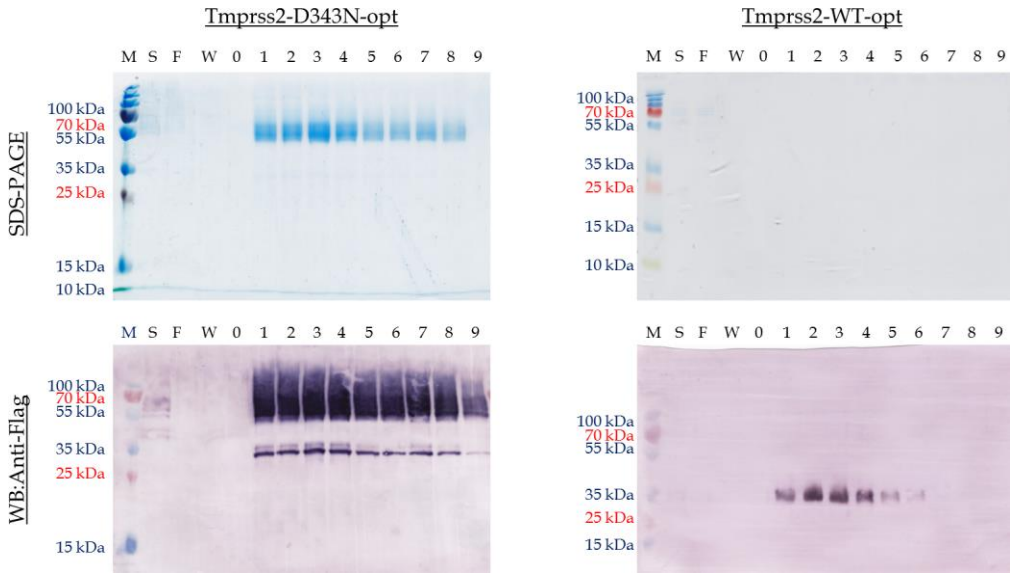


Figure 3.35. Affinity capture of 2 L Shake flask scale expression of Tmprss2 (optimized) from *P. pastoris*. SDS-PAGE and Western blot are shown with the harvested culture supernatant (S), the flow through of the Ni-NTA column (F), the wash fraction (W) and the elution fractions (0-9) of the Ni-NTA affinity capture. Tmprss2 was completely captured from the supernatant. For Tmprss2-D343N the elution fractions 1 to 8 were pooled for further analysis. For Tmprss2-WT, the protein was only visible on the Western blot. The elution fractions 1 to 5 were pooled. Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

3.2.4.3 Evaluation of human serum albumin as a cargo protein for Tmprss2

The low expression levels of Tmprss2 and its fast degradation in the medium during cultivation present major bottlenecks in its production. Human serum albumin (HSA) is expressed at very high levels in *P. pastoris*. Furthermore, it is highly stable in the culture supernatant, which allows for the accumulation of high amounts of the protein over an extended period of time. For this reason, it was tested whether the fusion of HSA to Tmprss2 presented a viable option to improve the expression yields of the latter protein in *P. pastoris* by two means. First, the strongly secreted HSA could aid the secretion of Tmprss2 and serve as a cargo protein. Secondly, the stability of HSA in the supernatant might also protect Tmprss2 from degradation. This would allow extending the methanol induction phase, i. e. the productive time of a single batch of previously grown cells. For this, HSA was fused N-terminally to Tmprss2 separated by a GSGG linker and a PreScission 3C protease recognition site. Hence, the secretion of the fusion protein HSA-Tmprss2 is mediated by the native HSA secretion signal.

The vector pPICZ-HSA was generated by replacing the original MF-α ss of pPICZα-A (Thermo Fisher Scientific), located upstream of the MCS, with the *HSA* gene (including the sequences of the

GGSG linker and the 3C site). HSA was cloned in frame with the c-myc and His6 tags of pPICZα-A, located downstream of the MCS, to enable individual expression of tagged HSA. The initial test expression of the HSA-Tmprss2 fusion protein was limited to Tmprss2-D343N, as the individual mutant form had previously been expressed at quantifiable levels (in contrast to Tmprss2-WT). The gene *Tmprss2-D343Nopt* was cloned into the MCS of pPICZ-HSA in frame with *HSA*. The resulting vector was named pPICZ-HSA-Tmprss2-D343Nopt. The cloning steps are described in detail in chapter 2.3.

The vectors pPICZ-HSA-Tmprss2-D343Nopt and pPICZ-HSA ("empty") were respectively linearized and transformed into *P. pastoris* KM71H. The transformed clones were designated by the prefixes "h" (KM71H/pPICZ-HSA) and "ht" (KM71H/pPICZ-HSA-Tmprss2-D343N). Following primary selection on YPDS Zeocin agar plates and single clone isolation, a high-throughput expression screen was performed in 500 µL BMMY cultures. The jackpot clone KM71H/pPICZα-Tmprss2-D343Nopt o108 was included in the experiment as a comparative control. Methanol induction was maintained for 48 hpi. Slot blots were performed with the culture supernatants (§ 2.9.4). In a first test round, 24 clones of KM71H/pPICZ-HSA-Tmprss2-D343N were tested. The overall expression levels were significantly lower than that of the jackpot clone o108. As a consequence, 68 additional clones were analyzed in the same way. Furthermore, 71 clones of KM71H/pPICZ-HSA were cultivated in parallel for comparison. The strongest clone from the initial set (ht11), the jackpot clone o108 and empty KM71H cells were included as controls. For the slot blots, the supernatants of the clones expressing HSA were diluted 1:10. The supernatants of the clones expressing HSA-Tmprss2-D343N were not diluted prior to the analysis. The results of both expression screens are shown in Figure 3.36. The expression of HSA-Tmprss2-D343N was significantly lower than the expression of the individual proteins (HSA or Tmprss2-D343N) in all tested clones.

To evaluate this observation, the total cell lysate and the supernatant of the maximum producer clone of HSA-Tmprss2-D343N (ht11) and a weaker producer (ht31) were analyzed by SDS-PAGE and Western blot (§ 2.9). The maximum producer clones of the individual constructs (HSA: clone h34; Tmprss2-D343N: o108) served as comparative controls. The expression experiment was repeated with these clones. The harvested culture supernatants were concentrated twenty-fold by TCA precipitation (§ 2.8.6). Cell extracts were generated (§ 2.8.4). The results of the SDS-PAGE and the Western blot are shown in Figure 3.37. Unlike the individual proteins (HSA or Tmprss2-D343N), the HSA-Tmprss2-D343N fusion protein showed a strong smeared signal in the insoluble cell extract fraction. One possible explanation could be that the fusion protein cannot be secreted properly. Speculatively, HSA-Tmprss2-D343N might get stuck in the ER, where it would be subsequently degraded.

All in all, the N-terminal fusion of HSA to Tmprss2 did not lead to improved expression yields of the target protein. For this reason, this project path was not pursued any further in this PhD thesis. Instead, it was tested, whether the expression process could be optimized in order to improve the product yields of Tmprss2 with the current jackpot clones.

Figure 3.36. Slot blots of the expression screen of HSA and HSA-Tmprss2-D343N.
In a first test, 24 clones of KM71H/pPICZ-HSA-Tmprss2-D343Nopt (ht1-h24) were analyzed for the expression of the fusion protein HSA-Tmprss2-D343N. Subsequently, 68 more clones (ht25-ht92) were analyzed. In parallel, 71 clones of KM71H/pPICZ-HSA (h1-h71) were analyzed in comparison. Controls are denoted in bold-italics. The controls comprised purified Tmprss2-D343N as a positive control (PC). Moreover, culture supernatant the clones KM71H/pPICZα-Tmprss2-D343Nopt clone o108 and untransformed KM71H cells (negative control, NC) served as comparative controls. For each sample of KM71H/pPICZ-HSA-Tmprss2-D343Nopt and for the controls, 150 µL of culture supernatant were directly loaded per slot. For KM71H/pPICZ-HSA the supernatant was diluted 1:10 in PBS to a final volume of 150 µL. The slot blots were stained with a primary mouse α-His mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. The clones expressing the HSA-Tmprss2 fusion construct performed significantly weaker than the clones expressing the two individual proteins. The strongest clone expressing HSA-Tmprss2-D343N (ht11) is marked in bold and indicated by the black arrow.

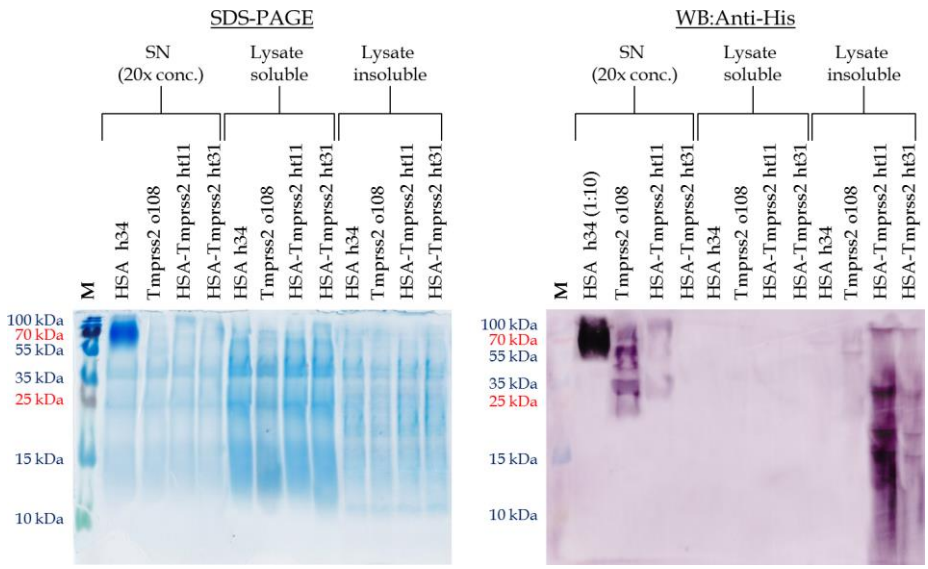


Figure 3.37. Analysis of the localization of the HSA-Tmprss2-D343N fusion protein in *P. pastoris*.
The culture supernatants and the cell extracts of the HSA expressing clone h34, the Tmprss2-D343N expressing clone o108 and the two HSA-Tmprss2-D343N expressing clones ht11 and ht31 were analyzed. The SDS-PAGE and Western blot are shown with the TCA precipitated supernatant (SN) (1:20 conc.) as well as the soluble and insoluble cell lysate fractions. The dominant signal of the HSA-Tmprss2-D343N fusion protein is presented by a strong smear in the insoluble extract fraction, which indicates proteolytic degradation. Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-His mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

3.2.5 Optimization of cultivation and downstream processing for the production of Tmprss2 in *P. pastoris*

Even after the gene optimization, Tmprss2 was only produced at low levels in *P. pastoris* from jackpot clones identified by extensive expression screens. Hence, it is unlikely that the expression yields could be sufficiently improved from the "gene / construct side". Particularly Tmprss2-WT was problematic in this regard, because it was hypothetically diminishing its own maximal production rate due to its enzymatic activity. Consequently, it could be impossible to generate producer clones with significantly higher expression yields.

Apart from the productivity of the generated producer clones, the cultivation strategy and the downstream processing are major factors that influence the effective space-time yields of the target protein. For this reason, it was assessed, whether minimizing the required process time and the loss of the target protein during the downstream process could improve the effective product yields of Tmprss2 with the current jackpot clones.

3.2.5.1 Direct capture of Tmprss2 from the culture supernatant using Q-Sepharose

In the previously described experiments, Tmprss2 was purified from harvested *P. pastoris* culture supernatant in two separate steps. The first step comprised diafiltration and concentration of the culture supernatant. The second step consisted of the His tag based affinity capture of the target

protein. This could result in a time-consuming process, particularly for large scale productions of Tmprss2. The amount of required time could be detrimental for the quality of the target protein. Moreover, each separate step during downstream processing is generally associated with a loss of the target protein. For this reason, it was analyzed whether the downstream processing could be streamlined.

One possible approach was to initially capture Tmprss2 directly from the culture supernatant using ion exchange chromatography to omit the diafiltration step. As a secondary effect, this would also result in a "pre-purification" of the target protein prior to further tag based affinity capture steps. This "double purification" would beneficially result in lower amounts of possible contaminant proteins in the final eluate. In combination with an ÄKTApilot (GE Healthcare), which is particularly suited for large sample volumes and high flow rates, this method would result in a time-saving process to purify Tmprss2. Furthermore, the ÄKTApilot can be operated with two separate columns. Hence, the eluent fraction of the ion exchange column (containing Tmprss2) could be directly and automatically loaded onto a second column for His tag based capture in the same run. This would minimize the process time and reduce the risk of protein degradation occurring over time.

To initially test the capture of Tmprss2 by ion exchange, Tmprss2-D343N was produced from clone o108 (KM71H/pPICZα-Tmprss2-D343N) in a Labfors bioreactor (§ 2.8.1). The culture supernatant was diluted 1:5 in in 25 mM Tris buffer, pH 9.0 to adjust the pH. The ion exchange capture of Tmprss2-D343N (estimated pI 6.43) from the diluted supernatant was successfully performed with a Q-Sepharose FF column (GE Healthcare) (§ 2.8.11). Tmprss2-D343N was eluted from the resin at a NaCl concentration of 500 mM (Figure 3.38A). Tmprss2 was subsequently captured from the respective ion exchange eluent fraction using a 1 mL HisTALON resin (GE) (Figure 3.38B). This shows that Q-Sepharose columns present an adequate method to directly capture Tmprss2 from the culture supernatant. Furthermore, Tmprss2 can be further purified by His tag based affinity capture from the respective eluent fraction.

This could provide a time-saving, one-step setup using the ÄKTApilot. The proposed method would be particularly beneficial for large culture volumes, which could be essentially required to produce sufficient amounts of Tmprss2 for structural analyses.

An alternative approach to the use of two separate resins would be the direct tag based affinity capture of Tmprss2 from the culture supernatant. This method would take a pass on the benefits of the "double purification" described above. However, it would enable the fast and direct capture of Tmprss2 in one single purification step even without an ÄKTApilot. This would be particularly advantageous for smaller culture volumes (e. g. test expressions in shake flasks).

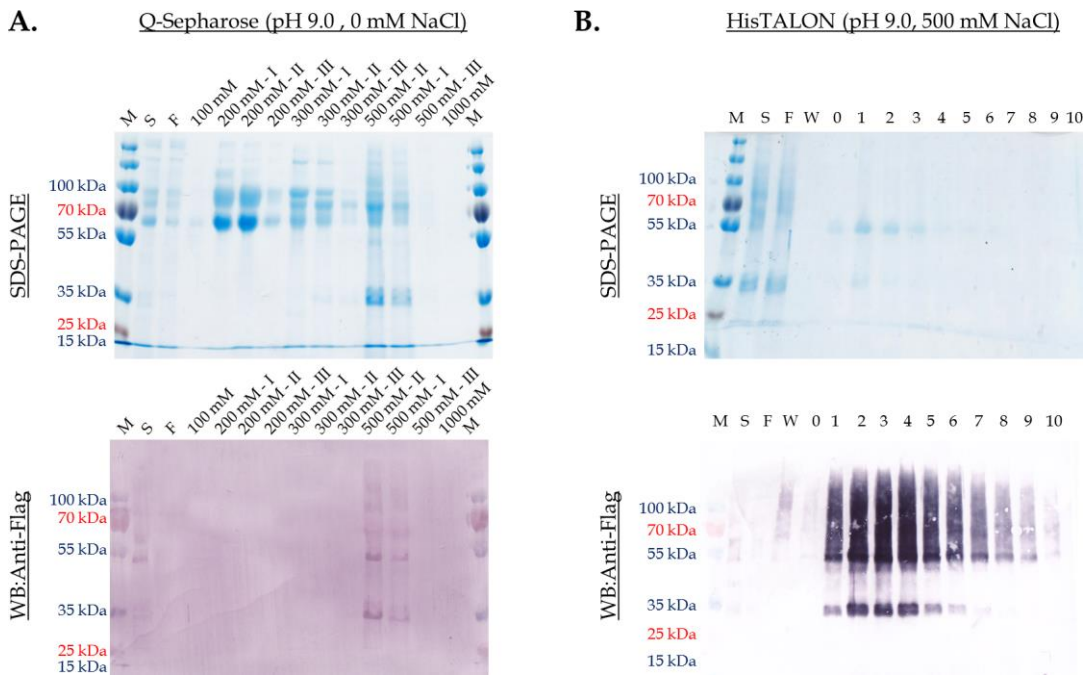


Figure 3.38. Direct capture of Tmprss2-D343N from *P. pastoris* culture supernatant.
At pH 9.0, Tmprss2-D343N was binding to Q-Sepharose FF up to a NaCl concentration of 500 mM (A). Subsequently, the protein could be purified from the eluent fraction (500 mM II, III) on a HisTALON resin (B). Indices: S: Supernatant; F: Flowthrough; 100 mM – 1000 mM: NaCl concentrations for Q-Sepharose elution; I-III: start, maximum and tail fractions of observed peaks following NaCl concentration increase; W: Wash fraction HisTALON; 0 to 10: eluent fractions HisTALON. Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

Hence, the possibility to perform the direct, tag based capture of Tmprss2 was assessed with the resin materials Ni-NTA HisTrap (GE Healthcare) or ANTI-FLAG (Sigma). For this, Tmprss2-D343N was expressed at 1 L scale in shake flask cultures from clone o108 (§ 2.8.1). The harvested culture supernatants were cleared by centrifugation and filtration as described in chapter 2.8.1, but not diafiltered. The experiment was performed with 20 mL culture supernatant samples. The detailed protocol is provided in chapter 2.8.10. Briefly, the binding to the materials was tested in batch. The samples were supplemented with 50 mM of binding buffer to adjust the pH (pH 7.4) for proper conditions of the target protein to the respective resin materials.

The direct capture on Ni-NTA material was tested with or without the addition of 1 mM $NiSO_4$ as a protecting supplement against chelating substances. Furthermore, the Ni-NTA samples were supplemented with 10 mM imidazole to provide more specific binding of the target protein to the resin. The ANTI-FLAG samples were not supplemented with any additional substances. However, two different conditions were tested for the elution of Tmprss2-D343N from the ANTI-FLAG resin: The elution at low pH (pH 3.5) and the competitive elution (against Flag peptide). Samples of the respective supernatants before and after binding ("non-bound"), the wash fractions and the eluent fractions were analyzed by SDS-PAGE and Western blot (§ 2.9.3). To evaluate the presence of left-over or precipitated target protein on the material, the used resin materials were each boiled

in 2x concentrated SDS loading buffer. The results are depicted in Figure 3.39. Ni-NTA binding was improved significantly by the addition of 1 mM $NiSO_4$ to the supernatant. The capture by ANTI-FLAG material resulted in protein degradation under both applied elution conditions. Particularly low pH elution was detrimental for the protein quality. It was concluded that the direct capture Tmprss2 from the culture supernatant by Ni-NTA under the addition of 1 mM $NiSO_4$ presented the most feasible method among the four tested alternatives. Therein, the use of Ni-NTA beneficially also comprises the more cost-efficient approach in comparison to ANTI-FLAG material.

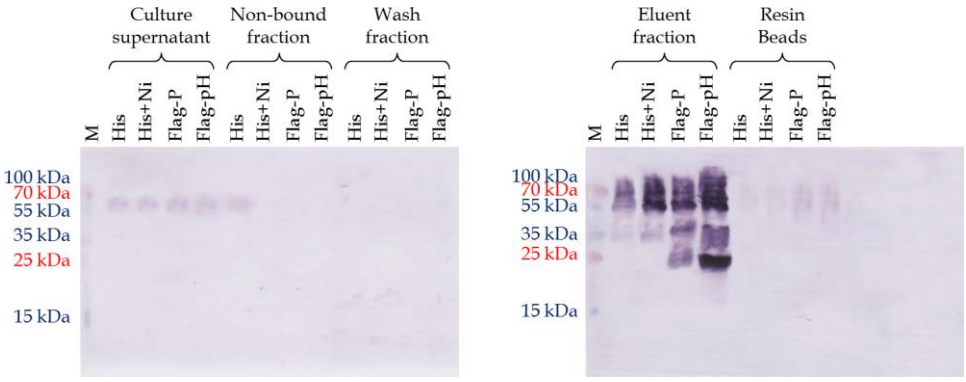


Figure 3.39. Binding test of Tmprss2 to Ni-NTA or ANTI-FLAG resin from *P. pastoris* culture supernatant. The direct binding of Tmprss2-D343N was tested with Ni-NTA and ANTI-FLAG resin. For Ni-NTA, the binding to the material was tested without (His) and with the addition of 1 mM $NiSO_4$ (His+Ni). The elution was performed competitively with imidazole. For ANTI-FLAG, the elution was performed either by low pH (Flag-pH) or by competition with Flag peptide (Flag-P). The samples comprise the (supplemented) culture supernatant before binding, the supernatant fraction after binding (non-bound), the wash fraction, the eluent fraction and the beads. Direct capture by Ni-NTA was significantly more efficient in the presence of 1 mM $NiSO_4$. Direct binding via ANTI-FLAG resulted in protein degradation. Staining of the SDS gel was performed with InstantBlue. The Western blot was performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

Taken together, the results of the two experiments show that the downstream process can be streamlined by directly capturing Tmprss2 from the culture supernatant in a time-efficient, one-step purification process. On the one hand, the direct His tag based capture of Tmprss2 presents a fast and cost-efficient method. On the other hand, the use of an ÄKTApilot would be particularly beneficial to handle large culture volumes during scale-up, as two columns could be managed in a single run. In this case, the "double purification" of Tmprss2 using the Q-Sepharose ion exchange column prior to His tag based capture would present an appropriate approach, as it would conceivably reduce the risk of contaminant proteins in the final eluate.

3.2.5.2 Iterative methanol induction to optimize the production process of Tmprss2

The production process with mut^S strains like KM71H and the *AOX1* promoter usually comprises two separate phases: The accumulation of cell mass under glycerol conditions and the methanol-induced expression phase, in which the cells do not proliferate at high rates anymore. The first phase usually takes up about 48 h. This is half of the time of the total production process of Tmprss2, as the methanol induction is equally maintained for 48 hpi to ensure the protein quality (§ 3.2.1). Consequently, each production process is "void" for half of the time, in which sufficient amounts of the producer cells have to be generated. As discussed in chapter 3.2.4.3 for the HSA fusion protein, the prolonged expression from one generated batch of cells might present a viable option to enhance the space-time yields of Tmprss2. Since the N-terminal fusion of HSA did not provide such a way, the iterative induction of the same batch of cells was assessed as an alternative approach to reach this goal. The idea was to harvest the supernatant at 48 hpi (as before). However, instead of discarding the cells, the pellet would be resuspended in fresh expression medium for a second round of expression (again for 48 hpi). Therein, it was of interest whether the iterative production phases would result in comparable expression yields of Tmprss2.

Additionally, the used media in this experiment were supplemented with of the antifoam (AF) J673A (Struktol). J673A had been reported to significantly enhance the total yield of GFP produced in *P. pastoris* as well as its secretion to the culture supernatant (Routledge et al. 2011). Hence, it was tested whether the addition of J673A was beneficial for the secretory production of Tmprss2 and enable higher yields in shake flask cultures. The experiment was initially limited to the the codon optimized jackpot clone KM71H/pPICZα-Tmprss2-D343Nopt o108, because only Tmprss2-D343N had previously been expressed at quantifiable (and thereby, comparable) levels (§ 3.2.4.2). The expression was conducted at 1 L scale in shake flask cultures (§ 2.8.1). The cells were grown in 5 L of BMGY medium with 1 % (v/v) J673A (BMGY+AF) for 48 h. The grown cells were harvested and resuspended in 1 L BMMY. An OD_{595} of 80 was measured for the expression culture. Directly after the OD_{595} measurement, 1 % (v/v) J673A was added to the culture medium (BMMY+AF) and the production phase was started.

At 48 hpi, the culture supernatant was separated from the cells by centrifugation at 3.000 x g. The harvested cells were resuspended in 1 L of fresh BMMY medium for the second (iterative) production phase. An OD_{595} of 86 was measured, showing that the cells did not proliferate significantly since the first methanol induction of the culture. This was anticipated due to KM71H being a mut^S strain. Thus, the ratio of the nutrients (medium) to the cells in the second production phase was comparable to the first production phase. J673A was added again to the medium (BMMY+AF) and the cultivation was continued. The expression was again conducted for 48 hpi, after which the supernatant was harvested as before.

Tmprss2 was directly captured on Ni-NTA from the respectively harvested and cleared culture supernatants of the two iterative production phases using the previously developed protocol (§ 3.2.5.1). As described in chapter 3.2.5.1, the pH of the supernatants was adjusted to pH 7.4 through the addition of an appropriate sodium phosphate buffer. The supernatants were supplemented with 10 mM imidazole and 1 mM $NiSO_4$. The supplemented supernatants were then respectively loaded onto a 1 mL Ni-NTA column to capture Tmprss2-D343N (§ 2.8.11). The fractions of the purification process were analyzed by SDS-PAGE and Western blot (Figure 3.40).

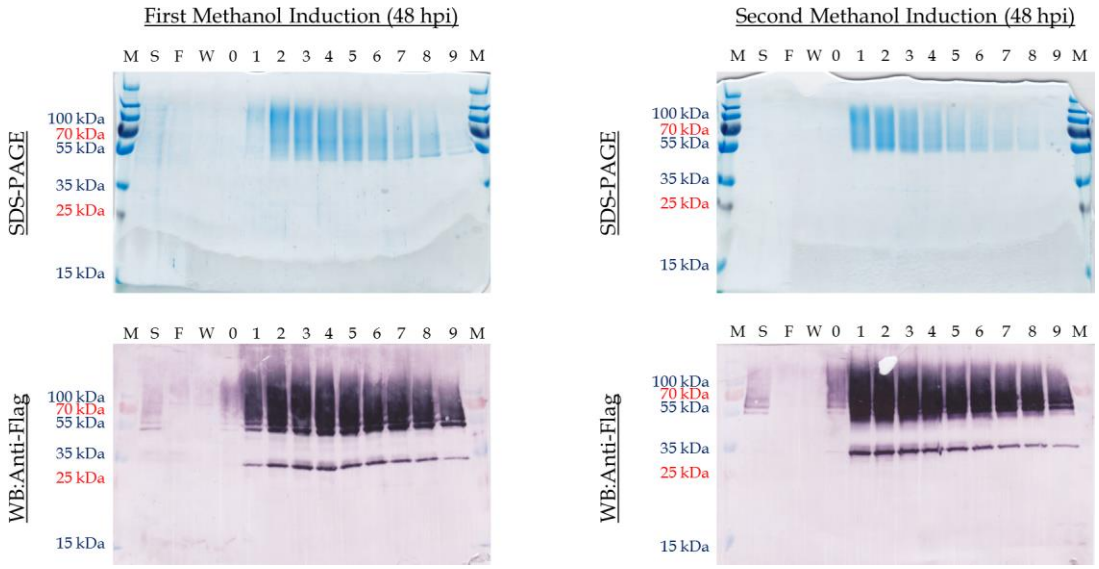


Figure 3.40. Direct Ni-NTA affinity capture of Tmprss2-D343N from 1 L shake flask cultures in *P. pastoris*. The expression of Tmprss2-D343N from clone o108 was analyzed in a 1 L expression culture. The cells were used for two subsequent induction phases à 48 hpi after their initial growth. The Ni-NTA affinity capture was performed directly from the culture supernatant in the presence of 1 mM $NiSO_4$. SDS-PAGE and Western blot are shown with the harvested culture supernatant (S), the flowthrough (F), the wash fraction (W) and the elution fractions (0 to 9). Tmprss2 was virtually completely captured from the supernatant in both cases. Both induction phases resulted in the expression of similar amounts of Tmprss2-D343N. Staining of the SDS gels was performed with InstantBlue. The Western blots were conducted with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

The elution fractions 2 to 8 (first induction) and 1 to 8 (second induction) were pooled and dialyzed (§ 2.8.7). The protein concentration was measured with a Nanodrop (§ 2.6.14). Tmprss2-D343N was expressed at similar levels of 0.6 mg/L and 0.7 mg/L during the two production phases. This proves that the iterative induction of the cells is possible and results in comparable expression yields. Furthermore, the addition of J673A to the culture media resulted in increased expression yields of Tmprss2-D343N. The respective yields in both production phases were considerably improved in comparison to the previously described experiments with the same codon optimized jackpot clone (o108) (§ 3.2.4.2).

Consequently, it was tested, if the optimized protocol would also improve the expression yield of Tmprss2-WT using the optimized jackpot clone KM71H-pPICZα-Tmprss2-WTopt o20. For this, the expression was conducted at an identical scale of (1 L of expression culture). The cells were

grown in 5 L BMGY+AF medium for 48 h. Two iterative rounds of production (48 hpi each) were performed in BMMY+AF medium with the same batch of cells. As described above, the OD_{595} of the expression culture was measured at 0 hpi. Thereby, an OD_{595} of 71 and an OD_{595} of 78 were measured for the first and the second induction phase, respectively. Tmprss2-WT was captured directly from the harvested and cleared culture supernatant on a 1 mL Ni-NTA column. The results are depicted in Figure 3.41. The elution fractions were pooled, concentrated to 1 mL with a Vivaspin 6 (10 kDa) column and dialyzed (§ 2.8.7, 2.8.8). The protein concentration was measured with a Nanodrop (§ 2.6.14). The optimized protocol resulted in the production of quantifiable amounts of Tmprss2-WT. The expression was quantified at 70 µg/L (first induction) and 80 µg/L (second induction). Thus, it was finally possible to correlate the expression levels of Tmprss2-WT and Tmprss2-D343N from the respective jackpot clones o20 (WT) and o108 (D343N). Therein, the expression level of Tmprss2-WT is around ten-fold lower in comparison to Tmprss-D343N.

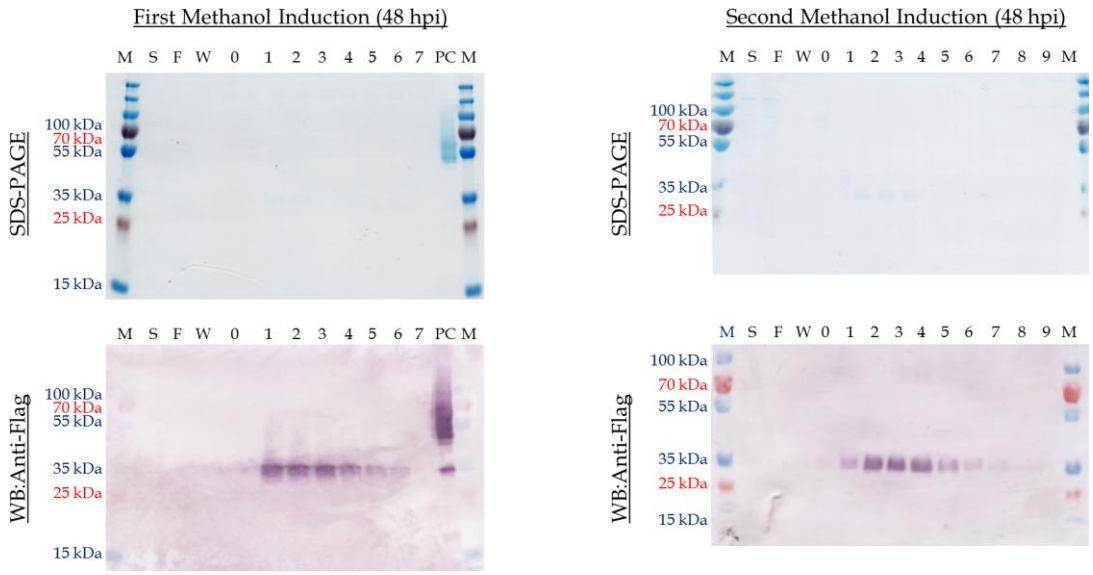


Figure 3.41. Direct Ni-NTA affinity capture of Tmprss2-WT from 1 L shake flask cultures in *P. pastoris*.
The expression of Tmprss2-WT from clone o20 was analyzed in a 1 L expression culture. The cells were used for two subsequent induction phases à 48 hpi after their initial growth. The Ni-NTA affinity capture was performed directly from the culture supernatant in the presence of 1 mM $NiSO_4$. SDS-PAGE and Western blot are shown with the harvested culture supernatant (S), the flowthrough (F), the wash fraction (W) and the elution fractions (0 to 9). Tmprss2-WT was completely captured from the supernatant in both cases. Tmprss2-D343N served as a positive control (PC). Both induction steps resulted in similar amounts of Tmprss2-WT. Staining of the SDS gels was performed with InstantBlue. The Western blots were performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

In summary, the volumetric expression yields of Tmprss2 were significantly improved by the overall optimized expression and purification process. The yield of Tmprss2-D343N was improved by a factor of 4x compared to the initially used protocol with the original gene, with which 0.15 mg/L were obtained from 1 L shake flask expression cultures (§ 3.2.2). Moreover, Tmprss2-WT was expressed at quantifiable amounts with this enhanced method. The cultivation of the cells under the addition of the AF reagent J673A resulted in higher yields of the secreted target protein.

Furthermore, the direct capture prevented the loss of the target protein due to additional steps, such as diafiltration and concentration. In addition to this, the iterative methanol induction resulted in comparable productivity and thus presents a considerably time-saving and work-efficient improvement over the original process.

3.2.5.3 Evaluation of long-term storage possibilities for Tmprss2

For crystallization, large amounts of highly pure protein are required. Due to its low expression levels, Tmprss2 requires the accumulation of the necessary amounts of protein over time in a series of production batches. However, Tmprss2 displays only low stability. During cultivation, first signs of degradation are evident from 72 hpi onwards (§ 3.2.1.1). During storage at 4 °C, degradation of the affinity captured protein was observed after approximately two weeks. This is visibly evident in the Western blot positive controls comprising purified Tmprss2-D343N, which were used during the test expression in insect cells or the HSA fusion tests (§ 3.2.3.2 and 3.2.4.3). Therefore, several storage conditions were investigated in order to identify a condition for long-term storage of Tmprss2 with relatively minimal loss of quality. The tested conditions included temperatures of -20 °C and -80 °C and different concentrations of glycerol (0, 25 or 50 %). The test was conducted for 28 days of storage with 15 ng samples of the affinity captured Tmprss2-D343N (§ 3.2.5.2). An additional control sample stored at 4 °C. The samples were analyzed by SDS-PAGE and Western blot (§ 2.9). The results are shown in Figure 3.42. No visible degradation was observed for all samples stored at -20 °C and -80 °C. In contrast, significant degradation was evident for the sample stored at 4 °C. Consequently, the storage of Tmprss2 at either -20 °C or -80 °C presents an appropriate alternative to storage at 4 °C, regardless of the added amount of glycerol.

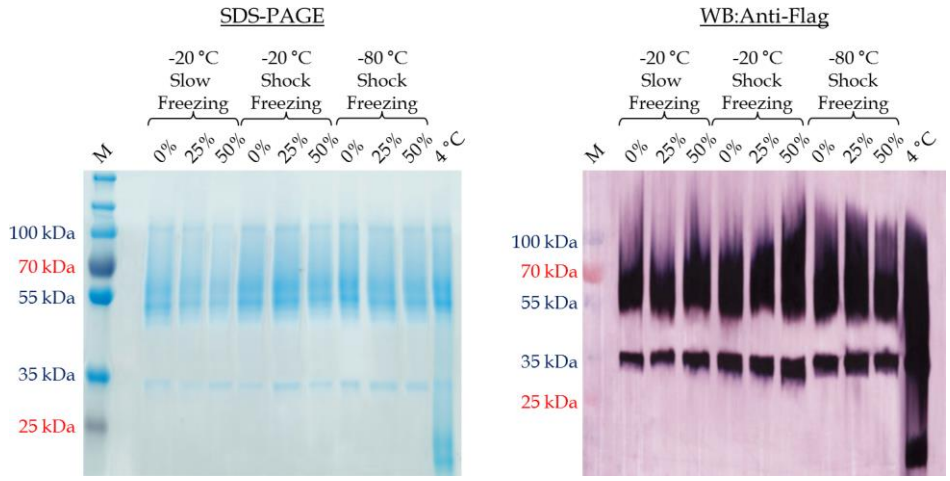


Figure 3.42: Long-term storage stability test of Tmprss2-D343N.
The storage was tested at -20 °C or -80 °C for 28 days with 15 ng samples of Tmprss2-D343N, which were supplemented with 0, 25 or 50 % of glycerol. The samples were either frozen slowly in a Mr. Frosty™ Freezing Container with isopropanol (Thermo Fisher Scientific) ("slow freezing") or quickly frozen with liquid nitrogen ("shock freezing"). A sample stored at 4 °C served as a control. In contrast to the control, all frozen samples show significantly less degradation. Staining of the SDS gels was performed with InstantBlue. The Western blots were performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

3.2.6 Functional analysis of recombinant Tmprss2

A preliminary enzymatic *in vitro* assay with Tmprss2-WT and Tmprss2-D343N was conducted. Therein, the functionality / activity of Tmprss2-WT should be initially evaluated. Moreover, proteolytic activity of Tmprss2-WT for the cleavage of hemagglutinin (HA) H1 was compared to the activity of the mutant form Tmprss2-D343N.

Tmprss2 should cleave the unprocessed H1 (HA0) into two split halves (HA1 and HA2). This was tested *in vitro* using the previously produced recombinant Tmprss2-WT and Tmprss2-D343N proteins (§ 3.2.5.2). Recombinant hemagglutinin H1 produced in Hi5 cells was used as a substrate for the reaction, as described in chapter 2.9.6. Briefly, equal amounts of Tmprss2-D343N and Tmprss2-WT were respectively mixed with HA H1 (HA0) and incubated at 37 °C. An according mock control was set up without Tmprss2. Samples were taken after 3 h and 16 h of incubation and analyzed by Western blot (§ 2.9.3). The results are shown in Figure 3.43. Tmprss2-WT completely cleaved H1 after 3 h, while Tmprss2-D343N showed only marginal activity at this time. After 16 h, slight cleavage was also observed for the samples with Tmprss2-D343N, albeit less than the amount of cleavage caused by Tmprss2-WT after 3 h. For the mock control, no specific cleavage was visible after 3 h or 16 h of incubation. Furthermore, protein degradation of H1 was visible after the extended incubation at 37 °C for 16 h.

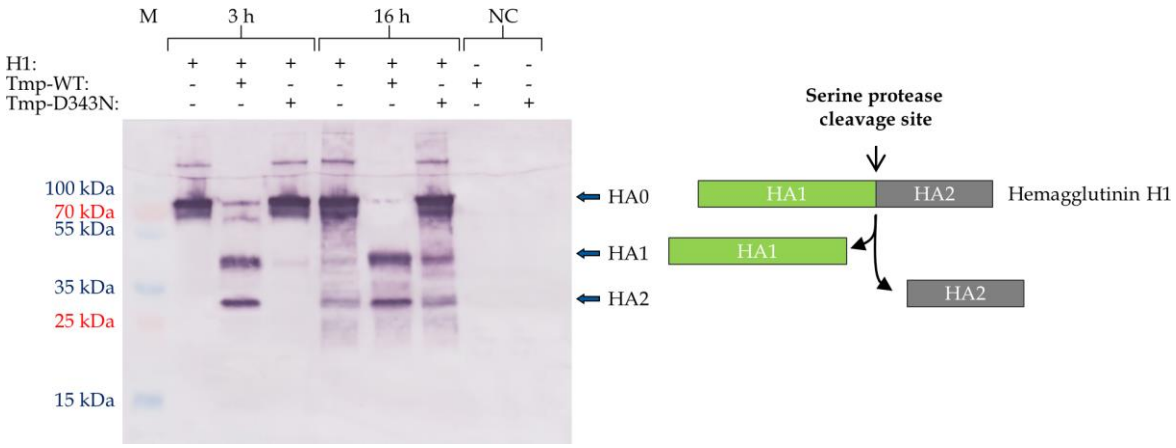


Figure 3.43. Enzymatic cleavage of hemagglutinin H1 by recombinant Tmprss2-WT and Tmprss2-D343N. For an initial functional analysis, the ability of Tmprss2-WT (Tmp-WT) and Tmprss2-D343N (Tmp-D343N) to enzymatically cleave the viral substrate hemagglutinin H1 (H1) from its unprocessed form (HA0) into the subdomains HA1 and HA2 was tested *in vitro*. For the reaction, 10 µg of H1 were each mixed with 10 ng of either Tmprss2 form and incubated at 37 °C. As a mock control, H1 was incubated without Tmprss2. Samples were taken after 3 h and 16 h of incubation. Furthermore, Western blot negative controls (NC) were set up comprising only Tmprss2-D343N or Tmprss2-WT without H1. The HA1 and HA2 subdomains are indicated by the respective arrows. The Western blot was performed with a primary rabbit α-PR8HA (H1) pAb and a secondary goat AP-conjugated α-rabbit pAb. Colorimetric staining was done with NBT and BCIP. Mw standard (M): PageRuler Prestained Plus.

The enzymatic cleavage of hemagglutinin H1 by the produced and purified recombinant Tmprss2 could be shown in this experiment. Tmprss2-WT displays a specific enzymatic activity on its viral substrate H1. The activity exerted by Tmprss2-WT is significantly higher than that of the mutant

form D343N. This is congruent with the previously observed reduced rate of zymogenic cleavage by Tmprss2-D343N, which could be significantly accelerated by the addition of Tmprss2-WT (§ 3.2.1 and 2.3). Furthermore, both forms of Tmprss2 could (at different rates) carry out the same defined cleavage of H1. This indicates that the mutant form D343N can still specifically bind to H1. However, this experiment does not allow distinguishing between the processed and the unprocessed protein molecules in the reaction mix. Hence, the presented results could not be used to analyze, whether Tmprss2-D343N has to be zymogenically processed in order to efficiently bind to the substrate H1.

All in all, Tmprss2 was identified as a difficult to express target protein in this work. Particularly Tmprss2-WT was produced at very low levels compared to the activity knockdown mutant Tmprss2-D343N. *P. pastoris* was shown to be a suitable expression system for the secretory production of Tmprss2. An optimized protocol for the expression of Tmprss2 was established using codon optimized jackpot clones (o20 and o108). The time-efficiency of the production process was substantially improved through iterative methanol induction of a single batch of grown cells. The downstream process was streamlined through fast, direct affinity capture of Tmprss2 from the harvested culture supernatant after the production phase. By this, the effective product yields were also significantly enhanced. Finally, it was proven that the recombinant Tmprss2 can enzymatically cleave H1 *in vitro*. Still, the question remains whether the activity of Tmprss2-D343N is sufficiently low for co-crystallization with HA H1.

4 Discussion

4.1 Evaluation of the established RMCE System in *P. pastoris*

Common transformation methods in *P. pastoris* rely on stable transformation by homologous recombination. This approach faces some drawbacks such as off-target integration of the expression vectors with unpredictable expression results. Moreover, the transformation generally is limited to the number of the currently characterized genomic loci for the integration of the GOI. Still, even if an expression vector is targeted to an established genomic locus, the resulting producer clones display very diverse expression yields. Hence, the generation and screening of high producer ("jackpot") clones usually presents a time-consuming and laborious bottleneck for the expression of recombinant target proteins. Minimizing the required effort and time-consumption for the generation of producer clones with uniform expression patterns is an important goal in order to streamline the expression process. Consequently, this work aimed at the establishment of a Flp/FRT RMCE system in *P. pastoris* to enable flexible and highly specific insertion of any GOI into a tagged genomic locus.

FRT F_3 impairs protein translation in *P. pastoris* in a location between promoter and GOI

The preliminary model system to establish the RMCE in *P. pastoris* relies on the constitutive *GAP* promoter and the integration of the tagging vectors into its genomic locus by homologous recombination. The initially devised RMCE cassette for *P. pastoris* was designed to be compatible to the multi-host expression system mHost-XS and its versatile multi-purpose vector pFlpBtM (Meyer et al. 2013). However, the combination "Promoter – F_3 – GOI", required for the RMCE cassette of pFlpBtM, impaired the protein expression of the GOI in *P. pastoris*. The influence was identified to occur at the level of translation.

Assumingly, F_3 is transcribed into the 5'-UTR of the mRNA of the GOI, where its secondary structure might hamper the initiation of protein translation. This hypothesis is supported by studies about the mRNA leader region, which report that both inhibitory and activating regions inside the 5'-UTR affect the translation efficiency (Dvir et al. 2013). This can either be related to secondary structures that influence ribosomal attachment or movement or to trans-acting factors, which bind to the 5'-UTR and influence the translation machinery (Mignone et al. 2002, Chatterjee and Pal 2009). In contrast to *P. pastoris*, no critical impact of FRT F_3 was observed in HEK293-6E, CHO LEC 3.2.8.1, SF21 and Hi5 cells so far (Meyer et al. 2013). All of these eukaryotic cell lines are significantly different from the methylotrophic yeast. Therefore, it is possible that the secondary structure of FRT F_3 acts as a critical element only in *P. pastoris*, for example by cross-interacting with an inhibiting 5'UTR trans-acting factor. Still, an in-depth analysis of this phenomenon would

have been beyond the scope of this thesis. For this reason, this matter was not further analyzed in favor of finding a working alternative to establish the RMCE system in *P. pastoris*.

The expression of the fluorescent model GOI is rescued by relocating FRT F_3 inside *P. pastoris* RMCE master cell lines

The expression could be fully rescued by relocating FRT F_3 upstream of the *GAP* promoter inside the genomic *GAP* locus (master cell line V5). Thereby, the initially intended compatibility to the mHost-XS was sacrificed. Consequently, the array of screenable expression hosts in the mHost-XS cannot be extended to *P. pastoris* using the RMCE system. Despite that, the identified working setup for the RMCE in *P. pastoris* offers a higher grade of flexibility. Not only the GOI, but also the promoter can be exchanged at will in *P. pastoris* within a single RMCE master cell line. This trait is beneficial, as the choice of the promoter can significantly influence the protein yields in *P. pastoris* depending on the target protein. For example, the usually weaker *GAP* promoter was shown to catch up to the expression levels of the *AOX1* promoter or to outperform it for some target proteins (Macauley-Patrick et al. 2005, Yang et al. 2015). Conceivably, this phenomenon may apply to many potential target proteins. The *P. pastoris* RMCE system hence is advantageous on the grounds of testing different promoters, including heterologous and even artificial promoters, without the need of generating a completely new RMCE master cell line every time.

On the other hand, the critical combination of "Promoter – F_3 – GOI", which inhibits the protein translation, could be purposefully applied to any gene of *P. pastoris* to perform a targeted expression / translation knockout. This approach could synergize particularly well with the Flp recombinase based marker gene regeneration in *P. pastoris* (Cregg and Madden 1989). This method is used to excise a stably integrated selection cassette through the recombination of an identical pair of FRT F_{WT} sites in parallel *cis* orientation (Cregg et al. 1989). Usually, the excision of the selection cassette leaves a "scar" (the recombined FRT F_{WT} site) in the respective genomic locus. If such a cassette would be inserted between a specifically targeted promoter and the downstream gene, the excision of the selection cassette would result in the "Promoter – FRT site – GOI" knockout combination.

The auxotrophic selection trap Δ*his4* is an efficient approach for clone selection in *P. pastoris*

During the generation of the *P. pastoris* RMCE master cell lines, a model for a novel kind of auxotrophic selection traps (Δ*his4*) was successfully established. It presents a less stressful, stringent and low-cost alternative compared to the costly antibiotic selection traps that are commonly applied in cell culture (Verhoeyen et al. 2001, Wilke et al. 2011, Meyer et al. 2013, Baser et al. 2016). Apart from the *P. pastoris* strain GS115 (*his4*), other strains with different amino acid biosynthesis knockouts exist and can be used for selection. For example, the strain

JC303 (*arg4 ura3 his4*) exhibits an arginine, a uracil and a histidine deficiency (Lin-Cereghino et al. 2001). These genes would probably be feasible to create further auxotrophic selection traps in *P. pastoris*. This could be used to insert multiple RMCE cassettes, flanked by different pairs of FRT sites into the genome, similarly to binary CHO RMCE master cell lines (Baser et al. 2016). It should be noted that the principle of an auxotrophic selection trap is not tied to the RMCE system. The setup would be equally feasible for the integration of vectors using one FRT site, comparable to the Flp-In™ system (Thermo Fisher Scientific) used in mammalian cells.

Alternatively, auxotrophic selection traps could also be complemented by homologous recombination in *P. pastoris*. Currently applied expression vectors like pPIC9k (Thermo Fisher Scientific) rely on the delivery of the full amino acid biosynthesis genes, including their promoters and transcriptional terminators. This increases the size of the vector and thereby the length of the integrating DNA fragment, which in turn negatively influences the integration efficiency (Näätsaari et al. 2012). Even more so, the auxotrophic markers present extended homologous regions. This can facilitate undesired off-target integrations of the vector. These undefined integration events are often indiscernible from the targeted integration of the vector in regard of the selectivity of the transformants. The principle of the selection trap could counter these drawbacks and lead to a higher genomic targeting specificity, while also decreasing the size of the expression vectors.

This proposed approach would particularly benefit from the ability to perform highly precise genomic editing, for example through the Crispr/Cas9 system, which was recently established for *P. pastoris* by Weninger et al. (2016). By specifically excising a genomic region - e. g. the promoter and the ATG start codon of the wildtype *His4* gene - a defined *his4* knockout strain could be generated. This strain would display the same selective properties as GS115 (which was instead generated by random mutation). The prototrophy of the defined knockout strain could be restored with an appropriately designed expression vector, which delivers the missing elements (promoter and ATG start codon) to complement the knocked-out *his4* gene. Hence, the partial or random insertion of the vector should not result in prototroph clones anymore. This should minimize the amount of false positive clones.

The generation of producer cell lines by RMCE with comparable expression levels is possible using the established *P. pastoris* RMCE master cell line V5

A proof of principle for the RMCE was delivered with master cell line V5 in a preliminary exchange experiment. Therein, the principal functionality of Flp recombinase in *P. pastoris* was anticipated, as it had already been used for marker gene recycling (Cregg and Madden 1989). In comparison, the RMCE system established for *P. pastoris* in this work is based on the recombination of two pairs of heterospecific FRT sites in *trans* orientation (respectively on the vector and in the genomic tagging cassette). Still, it remains to be evaluated, if and how the size of

the exchangeable cassette (due to different GOIs) would influence the efficiency of the RMCE reaction.

Interestingly, all exchanged clones produced similar levels of y-eGFP, as evident by their exhibited green fluorescence. This shows that producer clones with comparable expression results can be expected through the defined RMCE reaction. A similar observation was reported for binary CHO RMCE master cell lines by Baser et al., where the expression levels of two tagged loci were shown to be cumulative (Baser et al. 2016). This might rely on a notable aspect of the RMCE system in comparison to the "classical" system by homologous recombination: The RMCE system is principally restricted to harbor only one copy of the GOI. A lower gene copy number can be advantageous for some proteins like alkaline phytase (Yang et al. 2012). However, it can be frequently observed that higher copy numbers result in higher expression levels (Liu et al. 2014). This was also shown for the secretory expression of Tmprss2 in this work.

A random RMCE tagging approach might help to identify new, suitable genomic loci for recombinant protein expression, similar to the generation of CHO RMCE master cell lines (Baser et al. 2016). Once a favorable genomic locus is identified, the RMCE system enables the subsequent specific insertion of any GOI into it. Concomitantly, the established genomic loci like the *AOX1* or the *GAP* locus would be left in their innate state for the possible insertion of further GOI by homologous recombination. This dual approach could be used to enhance the gene copy numbers or to provide the expression of accessory proteins and chaperones to aid the expression of a GOI. Furthermore, a loxP site could be added to the RMCE cassette, as it is present in multi-host vector pFlpBtM-III (Steffen Meyer, HZI). This would allow for the subsequent insertion of additional vectors into an exchanged RMCE cassette by the cre-lox system (Sauer 1987).

As the RMCE master cell line harbors only one gene copy in a specifically tagged genomic locus, the expression of a GOI by different promoters should be comparable without any locus-dependent influences. This presents a promising approach to screen for the optimal promoter to produce a specific GOI. Moreover, high-throughput expression screens of gene construct libraries present an important task for recombinant protein expression, as truncated versions of a GOI can differ vastly in their expression yields and solubility (Bleckmann et al. 2016a). In *P. pastoris*, the RMCE system could provide an advantageous way for the expression screening of construct libraries due to the comparability of the different producer cell lines. A library of RMCE donor vectors could be transformed to perform an expression screen followed by the identification of the optimally produced gene construct by genomic PCR. This method could even enable subsequent transformations of the same construct by homologous recombination, if multi copy strains are desired. However, to establish a feasible construct library screening method, the efficiency of the

RMCE reaction needs to be optimized to achieve significantly higher numbers of positively exchanged clones.

All in all, the successfully established pilot RMCE system in *P. pastoris* delivers a proof of concept for the applicability of the system in the yeast and indicates that the generation of producer clones with predictable expression results is possible through a cassette exchange. It remains to be analyzed in-depth, whether the RMCE system provides a feasible, time-saving alternative to the well-established homologous recombination, particularly following random tagging, for the generation of suitable high producer clones of different GOI with reliable expression results.

4.2 Evaluation of the process optimizations for the expression of mouse Tmprss2

Tmprss2 is a highly interesting drug target playing a crucial role for influenza A infectivity. In this work, the secretory expression of the ectodomain of mouse Tmprss2 for structural and functional analyses was described. *P. pastoris* was identified as a suitable expression host. A protocol for the production of the active (wildtype) form and an activity knockdown mutant (D343N) of Tmprss2 was established and optimized to improve the expression yields as well as the downstream process.

***P. pastoris* is the most suitable expression host for the ectodomain of mouse Tmprss2**

The expression of Tmprss2 was initially tested in *P. pastoris*, because it is a frequently used expression system for the secretory production of other TTSPs like Hepsin, Matriptase and DESC1 (Friedrich et al. 2002, Somoza et al. 2003, Yuan et al. 2011). Thereby, Tmprss2 was shown to be a difficult to express target protein. During the initial test expression, it was observed that higher copy numbers of the vector resulted in higher yields of Tmprss2. The maximum producer ("jackpot") clones were selected on a Zeocin concentration of 1 mg/mL. These clones showed significantly higher productivity than all other tested clones, yielding 0.15 mg/L for Tmprss2-D343N, but only unquantifiable (though detectable) amounts of Tmprss2-WT.

In addition to *P. pastoris*, the expression was alternatively tested in eukaryotic hosts HEK293-6E, SF21 and Hi5. However, the expression was not successful in any of the tested systems. The expression could also not be shown in *E. coli* in a preliminary expression experiment (data was not presented). Tmprss2 is a multi-domain glycoprotein which includes many disulfide bonds. Because of this, it was not surprising that the expression in *E. coli* failed. This is supported by the fact that until now no other reports mention the expression of the full ectodomain of any TTSP in *E. coli*. Only the truncated serine protease domain of Matriptase-1 was reported to have been expressed in this system (Désilets et al. 2006, Beliveau et al. 2009). The observed lack of expression of Tmprss2 from SF21 or Hi5 insect cells in this work presents a notable difference to the (albeit unrelated)

Drosophila S2 cells, which had been used for the secretory expression of the ectodomains of Matriptase-2, Hepsin and DESC1 (Beliveau et al. 2009).

The deviating expression yields of Tmprss2-WT and Tmprss2-D343N in *P. pastoris* are presumably caused by the different proteolytic activities

Since all alternatively tested hosts did not produce any quantifiable amounts of soluble target protein, the production of Tmprss2 was continued in *P. pastoris*. Therein, several attempts were made to improve the expression yields. The codon optimization (GenScript) of the gene increased the expression yields by ca. two-fold. Through optimized cultivation conditions and a streamlined downstream processing, ~0.8 mg/L were obtained in shake flask cultures from a Tmprss2-D343N jackpot clone with the codon optimized gene (o108). This presents a 500 % increase over the originally obtained 0.15 mg/L. In contrast, Tmprss2-WT, while finally quantifiable, was produced at an order of magnitude less (~80 µg/L) from the respective codon optimized jackpot clone (o20).

This difference in expression yields might be caused by the active protease Tmprss2-WT cross-interacting with the host in an unknown way that eventually inhibits its own production. Tmprss2-WT and Tmprss2-D343N only differ in a single amino acid mutation. In D343N, the aspartic acid of the catalytic triad was replaced by asparagine, which resulted in a diminished proteolytic activity of the target protein. For this reason, it is plausible that the lower expression levels of Tmprss2-WT are not caused by limitations in protein processing or folding, but instead by its significantly higher activity compared to the mutant D343N. This is supported by a similar observation made by Zhao et al., who reported the individual expression of the serine protease domain of Matriptase-2 in *P. pastoris*. Therein, the active form, carrying one mutation (N164Q) to remove an N-glycosylation site, was expressed ten-fold lower than its inactive counterpart, in which the serine of the catalytic triad had been mutated (N164Q/S195A) (Zhao et al. 2013). Although the cause of this phenomenon was not identified, the authors speculated that the proteolytic activity of the active Matriptase-2 could be toxic to *P. pastoris* or exert an inhibitory effect on its expression (Sauer 1987). This is congruent with the conclusions drawn in this work. Comparable deductions were also made for the expression of active proteases in *E. coli* (Kwon et al. 2011). It is conceivable that similar effects can occur in any expression system.

This hypothesis is further supported by the fact that the individual expression of the serine protease domain of Tmprss2 as a truncated, simpler construct did not improve the expression yields compared to the whole ectodomain. The higher complexity of the full ectodomain was most likely not a deciding factor for the low expression levels, but rather the aforementioned activity of Tmprss2-WT. Notably, the expression yields of the individual serine protease domain of Matriptase-2 were reported at ~1 mg/L for the active form (N164Q) and ~10 mg/L for the inactive mutant (N164Q/S195A) by Zhao et al. (2013). This is a significantly higher level than the one

obtained for the Tmprss2 constructs in this work. However, the expression of the serine protease domain of Tmprss2 was only tested with the original CDS of the mouse gene. In contrast, the reported Matriptase-2 constructs were codon optimized, which might have substantially enhanced the expression levels. Moreover, both Matriptase-2 constructs include the mutation N164Q, which removes an N-glycosylation site (Yuan et al. 2011). This mutation might also influence the expression of the protein considerably. As a consequence, a definitive comparison is not possible. In addition to this, none of the currently available publications offer comparable expression yields for the full length ectodomain of Matriptase-2 in *P. pastoris*. In contrast, the secretory expression of the full ectodomains of the TTSPs Matriptase-2, Hepsin and DESC1 was reported with yields of 50-100 µg/L in Drosophila S2 insect cells (Beliveau et al. 2009). These results cannot be directly related to the ones obtained in this work due to the experimental differences (including the used expression host). However, it is still notable that the yield ranges within the same order of magnitude of Tmprss2-WT in this work. In the end, both publications and the results of this work consistently show that the secretory expression of TTSPs is generally a challenging task, regardless of the expression system.

The N-terminal fusion of HSA to Tmprss2 did not enhance the expression yields

A different approach to improve the expression yields of the Tmprss2 constructs was the N-terminal fusion of the strongly expressed and secreted human serum albumin (HSA) to Tmprss2 to act as a stabilizing carrier protein. However, the fusion protein accumulated insolubly inside the cells and was degraded. In contrast, a similar approach for an N-terminal HSA-VEGF165b fusion protein resulted in successful expression in *P. pastoris* (Zhu et al. 2012). Still, since Tmprss2 is a particularly difficult to express target protein, it is possible that it presented the critical bottleneck for the proper secretory expression of the fusion protein. Additionally, it is noteworthy that the recently reported HSA fusion proteins produced in *P. pastoris* frequently rely on the C-terminal fusion of HSA to provide stability to the target protein (Dou et al. 2008, Huang et al. 2008, Gou et al. 2012, Lei et al. 2012). This constellation was not tested for Tmprss2 yet.

Tmprss2 cleaves hemagglutinin H1 while the mutant D343N exerts reduced activity levels

In a first activity test *in vitro* Tmprss2-WT and Tmprss2-D343N showed defined catalytic cleavage of hemagglutinin H1. Thereby, the single mutation in the catalytic triad of Tmprss2-D343N resulted in a significant reduction of the activity, but not in a complete knockout. This is in accordance to prior reports on the catalytic triad of serine proteases (Craik et al. 1987, Carter and Wells 1988). However, it remains to be analyzed whether the rest activity is problematic to perform co-crystallization with hemagglutinin H1. If a significantly stronger knockdown would be required, this could be achieved by singly mutating the serine of the catalytic triad to alanine (S439A in mouse Tmprss2) (Carter and Wells 1988). Apart from a S439A single mutant, a D343N/S439A

double mutant of Tmprss2 should also display considerably lower activity levels than Tmprss2-D343N (Carter and Wells 1988).

Usually, zymogens undergo conformational rearrangements following proteolytic activation (Khan and James 1998). Tmprss2-D343N is undergoing zymogenic maturation at a significantly reduced rate compared to the wildtype. The heterogeneous population of processed and unprocessed Tmprss2-D343N proteins (or any other activity knockdown mutant) could be a detrimental factor for the co-crystallization of Tmprss2-D343N with HA H1. In this work, Tmprss2-WT was shown to accelerate the zymogenic activation of Tmprss2-D343N. Hence, Tmprss2-WT could be used to cleave Tmprss2-D343N in order to provide the required homogeneously processed population of Tmprss2-D343N for structural biology.

The co-crystallization of the fully processed Tmprss2-D343N with HA H1 would further be of advantage for another reason. Tmprss2 was shown to be an essential key player for influenza A virus pathogenesis by cleaving hemagglutinin H1 in living mice (Hatesuer et al. 2013). Accordingly, Tmprss2$^{(-/-)}$ knockout mice were protected from H1N1 virus infection (Prielhofer et al. 2013). Interestingly, other TTSPs did not facilitate the cleavage of H1 *in vivo*, despite displaying several structural similarities (particularly in regard of the serine protease domain). Hence, it is conceivable that the N-terminal LDLA and SRCR domains of the Tmprss2 ectodomain might play an important role for the interaction of Tmprss2 with the viral substrate H1. Since Tmprss2 is usually active (and processed) *in vivo*, it consequently makes sense to use the equally processed form of Tmprss2-D343N for co-crystallization with H1.

In summary, through several optimizations to the culture process and the downstream processing, a protocol for the robust secretory expression of the ectodomain of Tmprss2 was developed in this thesis. Moreover, the proteolytic activity of the recombinant proteins Tmprss2-WT and Tmprss2-D343N were shown in a preliminary enzymatic test assay. The assay revealed that the wildtype form could cleave both hemagglutinin H1 and Tmprss2-D343N, while the mutant form D343N displayed reduced activity levels.

5 Conclusions and Outlook

In this work, the establishment of a Flp/FRT based RMCE system in *P. pastoris* was demonstrated. A model system was established in the genomic *GAP* locus and different variants were comparatively analyzed for the expression of the GOI. The RMCE tagging cassette was designed specifically for *P. pastoris*. In contrast to the present CHO RMCE master cell lines, the cassette harbors a Zeocin resistance cassette in addition to a fluorescent marker to provide a stringent primary selection of tagged clones. Furthermore, the amino acid biosynthesis based selection trap Δ*his4* was developed. Δ*his4* presents a model for a novel kind of auxotrophic selection traps, which comprise a cost-efficient alternative to the antibiotic based counterparts. The selection trap in *P. pastoris* might also be principally applicable to the existing, classical homologous recombination system in order to improve the stringency of auxotrophic selection. The use of further amino acid biosynthesis genes as selection traps should be possible. This would provide a viable approach for the stringent insertion of multiple GOI, for example for the production of multi protein complexes.

Unlike originally intended, the *P. pastoris* RMCE system is not compatible to the multi-host expression system and its versatile vector pFlpBtM that is frequently used at the Helmholtz PSPF. This is due to the critical influence on protein translation by the FRT F_3 site in a position between promoter and GOI, which resulted in the necessity to change the RMCE cassette by placing FRT F_3 upstream of the promoter. Still, the RMCE system might provide a time-saving approach to perform comparative expression screens of construct libraries or for the production of defined levels of a GOI in *P. pastoris*. It remains to be tested, how the RMCE system performs for the exchange of different GOI with variable lengths and whether the expression is sufficient for more complex proteins than the model GOI used in this work (y-eGFP). However, it should be noted that the RMCE system relies on the insertion of a single gene copy of the GOI. This might significantly limit the capabilities of this method in an expression host like *P. pastoris*, which frequently benefits from the presence of multiple tandem insertions of an expression vector. To identify novel, suitable genomic loci for stable protein expression at high levels, random tagging should be performed. For this, the use of newly identified strong *P. pastoris* promoters like the *GCW14* promoter in place of the *GAP* promoter could additionally help to improve the expression yields.

In the second project presented in this thesis, a method was developed for the secretory expression of the full ectodomain of mouse Tmprss2 in both its active wildtype form and a mutant form with reduced activity (D343N). Thereby, it was shown that Tmprss2 is a difficult to express protein, which could only be successfully produced in *P. pastoris* among several tested expression hosts. Moreover, the expression level of the active wildtype form was shown to be significantly lower than that of the mutant form. Through optimizations of the CDS of the gene, the cultivation parameters and the downstream processing, the product yields were significant improved.

The functional integrity of Tmprss2-WT was proven in a preliminary enzymatic activity test. Furthermore, the mutant D343N could be shown to display a reduced proteolytic activity. However, it would still be important to assess whether the reduced activity of Tmprss2-D343N is sufficient for co-crystallization with hemagglutinin H1. If this is not the case, then a stronger knockdown variant should be generated (e. g. S439A) to ensure that the substrate is not prematurely cleaved. A complete knockout of the protease activity of Tmprss2 could also provide an interesting approach for a comparative crystallization of Tmprss2 in its processed and its unprocessed proenzyme form to evaluate the expected conformational change. The necessary proteolytic processing of the inactive Tmprss2 mutant could be mediated by Tmprss2-WT *in vitro*.

In perspective, it is worthwhile to assess further optimizations of the production process. First, it should be tested if the transfer of the optimized protocol to the bioreactor scale significantly increases the protein yields. Apart from the reported iterative methanol induction, a perfusion reactor process could present a viable alternative to handle greater culture volumes in a continuous process. Secondly, in the same context, the C-terminal fusion of HSA to Tmprss2 should be evaluated as a method to provide stability to the target protein in the supernatant. This could in turn allow for extended expression times. This would simplify the process and enhance the effective space-time yields, as a generated batch of producer cells could be productively used for a longer period of time.

Moreover, it would be possible to further genetically engineer the existing *P. pastoris* "jackpot" clones for Tmprss2. In a recent publication, the co-expression of the endoglycosidase EndoT was reported to enhance the protein quality of GPCRs for structural biology through direct deglycosylation *in vivo* (Claes et al. 2016). Furthermore, the same publication showed that the induction of the unfolded protein response (UPR) system in *P. pastoris* and the resulting production of ER chaperones can also result in higher amounts of correctly folded target protein. These optimizations could conceivably be beneficial to the production of Tmprss2 in *P. pastoris* and the subsequent downstream process.

Lastly, it is also of interest to assess whether the optimized expression protocol for Tmprss2 could also be applied for other TTSPs like Tmprss4, which is also an interesting candidate for structural biology to add to the research on influenza infection at the HZI.

Bibliography

- Afar, D. E., I. Vivanco, R. S. Hubert, J. Kuo, E. Chen, D. C. Saffran, A. B. Raitano and A. Jakobovits (2001). "Catalytic cleavage of the androgen-regulated TMPRSS2 protease results in its secretion by prostate and prostate cancer epithelia." Cancer Res **61**(4): 1686-1692.
- Ahmad, M., M. Hirz, H. Pichler and H. Schwab (2014). "Protein expression in Pichia pastoris: recent achievements and perspectives for heterologous protein production." Appl Microbiol Biotechnol **98**(12): 5301-5317.
- Alberts, B., A. Johnson, A. Lewis, M. Raff, K. Roberts and P. Walter (2002). Molecular Biology of the Cell, Garland Science.
- Altmann, F., E. Staudacher, I. B. Wilson and L. Marz (1999). "Insect cells as hosts for the expression of recombinant glycoproteins." Glycoconj J **16**(2): 109-123.
- Antalis, T. M., M. S. Buzza, K. M. Hodge, J. D. Hooper and S. Netzel-Arnett (2010). "The cutting edge: membrane-anchored serine protease activities in the pericellular microenvironment." Biochem J **428**(3): 325-346.
- Aoki, H., M. Nazmul Ahsan and S. Watabe (2003). "Heterologous expression in Pichia pastoris and single-step purification of a cysteine proteinase from northern shrimp." Protein Expr Purif **31**(2): 213-221.
- Apweiler, R., H. Hermjakob and N. Sharon (1999). "On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database." Biochim Biophys Acta **1473**(1): 4-8.
- Aricescu, A. R., R. Assenberg, R. M. Bill, D. Busso, V. T. Chang, S. J. Davis, A. Dubrovsky, L. Gustafsson, K. Hedfalk, U. Heinemann, I. M. Jones, D. Ksiazek, C. Lang, K. Maskos, A. Messerschmidt, S. Macieira, Y. Peleg, A. Perrakis, A. Poterszman, G. Schneider, T. K. Sixma, J. L. Sussman, G. Sutton, N. Tarboureich, T. Zeev-Ben-Mordehai and E. Y. Jones (2006). "Eukaryotic expression: developments for structural proteomics." Acta Crystallogr D Biol Crystallogr **62**(Pt 10): 1114-1124.
- Arnau, J., C. Lauritzen, G. E. Petersen and J. Pedersen (2006). "Current strategies for the use of affinity tags and tag removal for the purification of recombinant proteins." Protein Expr Purif **48**(1): 1-13.
- Arnolds, K. L. and J. V. Spencer (2014). "CXCR4: a virus's best friend?" Infect Genet Evol **25**: 146-156.
- Ascacio-Martinez, J. A. and H. A. Barrera-Saldana (2004). "Production and secretion of biologically active recombinant canine growth hormone by Pichia pastoris." Gene **340**(2): 261-266.
- Aw, R. and K. M. Polizzi (2016). "Liquid PTVA: a faster and cheaper alternative for generating multi-copy clones in Pichia pastoris." Microb Cell Fact **15**: 29.
- Baneyx, F. and M. Mujacic (2004). "Recombinant protein folding and misfolding in Escherichia coli." Nat Biotechnol **22**(11): 1399-1408.
- Baser, B., J. Spehr, K. Bussow and J. van den Heuvel (2016). "A method for specifically targeting two independent genomic integration sites for co-expression of genes in CHO cells." Methods **95**: 3-12.
- Beliveau, F., A. Desilets and R. Leduc (2009). "Probing the substrate specificities of matriptase, matriptase-2, hepsin and DESC1 with internally quenched fluorescent peptides." FEBS J **276**(8): 2213-2226.
- Berger, E. G., H. Clausen and R. D. Cummings (2012). Glycotechnology, Springer Science & Business Media.
- Bertram, S., A. Heurich, H. Lavender, S. Gierer, S. Danisch, P. Perin, J. M. Lucas, P. S. Nelson, S. Pöhlmann and E. J. Soilleux (2012). "Influenza and SARS-Coronavirus Activating Proteases TMPRSS2 and HAT Are Expressed at Multiple Sites in Human Respiratory and Gastrointestinal Tracts." PLoS ONE **7**(4): e35876.
- Bill, R. M. (2015). "Recombinant protein subunit vaccine synthesis in microbes: a role for yeast?" J Pharm Pharmacol **67**(3): 319-328.
- Bleckmann, M., S. Schmelz, C. Schinkowski, A. Scrima and J. van den Heuvel (2016a). "Fast plasmid based protein expression analysis in insect cells using an automated SplitGFP screen." Biotechnol Bioeng **113**(9): 1975-1983.
- Bleckmann, M., M. Schürig, F. F. Chen, Z. Z. Yen, N. Lindemann, S. Meyer, J. Spehr and J. van den Heuvel (2016b). "Identification of Essential Genetic Baculoviral Elements for Recombinant Protein Expression by Transactivation in Sf21 Insect Cells." PLoS One **11**(3): e0149424.
- Böttcher-Friebertshauser, E., C. Freuer, F. Sielaff, S. Schmidt, M. Eickmann, J. Uhlendorff, T. Steinmetzer, H. D. Klenk and W. Garten (2010). "Cleavage of influenza virus hemagglutinin by airway proteases TMPRSS2 and HAT differs in subcellular localization and susceptibility to protease inhibitors." J Virol **84**(11): 5605-5614.
- Böttcher, E., C. Freuer, T. Steinmetzer, H. D. Klenk and W. Garten (2009). "MDCK cells that express proteases TMPRSS2 and HAT provide a cell system to propagate influenza viruses in the absence of trypsin and to study cleavage of HA and its inhibition." Vaccine **27**(45): 6324-6329.
- Böttcher, E., T. Matrosovich, M. Beyerle, H.-D. Klenk, W. Garten and M. Matrosovich (2006). "Proteolytic Activation of Influenza Viruses by Serine Proteases TMPRSS2 and HAT from Human Airway Epithelium." Journal of Virology **80**(19): 9896-9898.
- Bretthauer, R. K. and F. J. Castellino (1999). "Glycosylation of Pichia pastoris-derived proteins." Biotechnol Appl Biochem **30 (Pt 3)**: 193-200.

- Brooks, C. L., M. Morrison and M. Joanne Lemieux (2013). "Rapid expression screening of eukaryotic membrane proteins in Pichia pastoris." Protein Science : A Publication of the Protein Society **22**(4): 425-433.
- Buchanan, S. G. (2002). "Structural genomics: bridging functional genomics and structure-based drug design." Curr Opin Drug Discov Devel **5**(3): 367-381.
- Bugge, T. H., T. M. Antalis and Q. Wu (2009). "Type II Transmembrane Serine Proteases." The Journal of Biological Chemistry **284**(35): 23177-23181.
- Byrne, B. (2015). "Pichia pastoris as an expression host for membrane protein structural biology." Curr Opin Struct Biol **32**: 9-17.
- Cabantous, S., T. C. Terwilliger and G. S. Waldo (2005). "Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein." Nat Biotechnol **23**(1): 102-107.
- Carter, P. and J. A. Wells (1988). "Dissecting the catalytic triad of a serine protease." Nature **332**(6164): 564-568.
- Carter, P. J. (2006). "Potent antibody therapeutics by design." Nat Rev Immunol **6**(5): 343-357.
- Cereghino, J. L. and J. M. Cregg (2000). "Heterologous protein expression in the methylotrophic yeast Pichia pastoris." FEMS Microbiol Rev **24**(1): 45-66.
- Chalfie, M. (1995). "Green fluorescent protein." Photochem Photobiol **62**(4): 651-656.
- Chalfie, M., Y. Tu, G. Euskirchen, W. W. Ward and D. C. Prasher (1994). "Green fluorescent protein as a marker for gene expression." Science **263**(5148): 802-805.
- Chatterjee, S. and J. K. Pal (2009). "Role of 5'- and 3'-untranslated regions of mRNAs in human diseases." Biol Cell **101**(5): 251-262.
- Chen, Y. W., M. S. Lee, A. Lucht, F. P. Chou, W. Huang, T. C. Havighurst, K. Kim, J. K. Wang, T. M. Antalis, M. D. Johnson and C. Y. Lin (2010). "TMPRSS2, a serine protease expressed in the prostate on the apical surface of luminal epithelial cells and released into semen in prostasomes, is misregulated in prostate cancer cells." Am J Pathol **176**(6): 2986-2996.
- Claes, K., K. Vandewalle, B. Laukens, T. Laeremans, O. Vosters, I. Langer, M. Parmentier, J. Steyaert and N. Callewaert (2016). "Modular Integrated Secretory System Engineering in Pichia pastoris To Enhance G-Protein Coupled Receptor Expression." ACS Synth Biol.
- Cohen, S. N., A. C. Chang, H. W. Boyer and R. B. Helling (1973). "Construction of biologically functional bacterial plasmids in vitro." Proc Natl Acad Sci U S A **70**(11): 3240-3244.
- Cormack, B. P., G. Bertram, M. Egerton, N. A. Gow, S. Falkow and A. J. Brown (1997). "Yeast-enhanced green fluorescent protein (yEGFP): a reporter of gene expression in Candida albicans." Microbiology **143 (Pt 2)**: 303-311.
- Cormack, B. P., R. H. Valdivia and S. Falkow (1996). "FACS-optimized mutants of the green fluorescent protein (GFP)." Gene **173**(1 Spec No): 33-38.
- Corpet, F. (1988). "Multiple sequence alignment with hierarchical clustering." Nucleic Acids Research **16**(22): 10881-10890.
- Craik, C. S., S. Roczniak, C. Largman and W. J. Rutter (1987). "The catalytic role of the active site aspartic acid in serine proteases." Science **237**(4817): 909-913.
- Cregg, J. M. (1987). Genetics of Methylotrophic Yeasts. Microbial Growth on C1 Compounds: Proceedings of the 5th International Symposium. H. W. van Verseveld and J. A. Duine. Dordrecht, Springer Netherlands**:** 158-167.
- Cregg, J. M. (2007). "Introduction: distinctions between Pichia pastoris and other expression systems." Methods Mol Biol **389**: 1-10.
- Cregg, J. M., K. J. Barringer, A. Y. Hessler and K. R. Madden (1985). "Pichia pastoris as a host system for transformations." Mol Cell Biol **5**(12): 3376-3385.
- Cregg, J. M. and K. R. Madden (1989). "Use of site-specific recombination to regenerate selectable markers." Mol Gen Genet **219**(1-2): 320-323.
- Cregg, J. M., K. R. Madden, K. J. Barringer, G. P. Thill and C. A. Stillman (1989). "Functional characterization of the two alcohol oxidase genes from the yeast Pichia pastoris." Mol Cell Biol **9**(3): 1316-1323.
- Cregg, J. M., Madden, K.R. (1988). Development of the methylotrophic yeast, Pichia pastoris, as a host system for the production of foreign proteins. Developments in industrial microbiology **v. 29**.
- Dalton, A. C. and W. A. Barton (2014). "Over-expression of secreted proteins from mammalian cell lines." Protein Sci **23**(5): 517-525.
- Daramola, O., J. Stevenson, G. Dean, D. Hatton, G. Pettman, W. Holmes and R. Field (2014). "A high-yielding CHO transient system: coexpression of genes encoding EBNA-1 and GS enhances transient protein expression." Biotechnol Prog **30**(1): 132-141.
- Davis, T. R., T. J. Wickham, K. A. McKenna, R. R. Granados, M. L. Shuler and H. A. Wood (1993). "Comparative recombinant protein production of eight insect cell lines." In Vitro Cell Dev Biol Anim **29A**(5): 388-390.
- De Schutter, K., Y.-C. Lin, P. Tiels, A. Van Hecke, S. Glinka, J. Weber-Lehmann, P. Rouze, Y. Van de Peer and N. Callewaert (2009). "Genome sequence of the recombinant protein production host Pichia pastoris." Nat Biotech **27**(6): 561-566.
- Désilets, A., J. M. Longpre, M. E. Beaulieu and R. Leduc (2006). "Inhibition of human matriptase by eglin c variants." FEBS Lett **580**(9): 2227-2232.

- Dissmeyer, N. and A. Schnittger (2011). "Use of phospho-site substitutions to analyze the biological relevance of phosphorylation events in regulatory networks." Methods Mol Biol **779**: 93-138.
- Donaldson, S. H., A. Hirsh, D. C. Li, G. Holloway, J. Chao, R. C. Boucher and S. E. Gabriel (2002). "Regulation of the epithelial sodium channel by serine proteases in human airways." J Biol Chem **277**(10): 8338-8345.
- Dou, W. F., J. Y. Lei, L. F. Zhang, Z. H. Xu, Y. Chen and J. Jin (2008). "Expression, purification, and characterization of recombinant human serum albumin fusion protein with two human glucagon-like peptide-1 mutants in Pichia pastoris." Protein Expr Purif **61**(1): 45-49.
- Durocher, Y. and M. Butler (2009). "Expression systems for therapeutic glycoprotein production." Curr Opin Biotechnol **20**(6): 700-707.
- Durocher, Y., S. Perret and A. Kamen (2002). "High-level and high-throughput recombinant protein production by transient transfection of suspension-growing human 293-EBNA1 cells." Nucleic Acids Res **30**(2): E9.
- Dvir, S., L. Velten, E. Sharon, D. Zeevi, L. B. Carey, A. Weinberger and E. Segal (2013). "Deciphering the rules by which 5'-UTR sequences affect protein expression in yeast." Proc Natl Acad Sci U S A **110**(30): E2792-2801.
- Eckart, M. R. and C. M. Bussineau (1996). "Quality and authenticity of heterologous proteins synthesized in yeast." Curr Opin Biotechnol **7**(5): 525-530.
- Ellis, S. B., P. F. Brust, P. J. Koutz, A. F. Waters, M. M. Harpold and T. R. Gingeras (1985). "Isolation of alcohol oxidase and two other methanol regulatable genes from the yeast Pichia pastoris." Mol Cell Biol **5**(5): 1111-1121.
- Esposito, D. and D. K. Chatterjee (2006). "Enhancement of soluble protein expression through the use of fusion tags." Curr Opin Biotechnol **17**(4): 353-358.
- Farre, J. C. and S. Subramani (2004). "Peroxisome turnover by micropexophagy: an autophagy-related process." Trends Cell Biol **14**(9): 515-523.
- Feng, Y., C. C. Broder, P. E. Kennedy and E. A. Berger (1996). "HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor." Science **272**(5263): 872-877.
- Francastel, C., M. C. Walters, M. Groudine and D. I. K. Martin (1999). "A Functional Enhancer Suppresses Silencing of a Transgene and Prevents Its Localization Close to Centromeric Heterochromatin." Cell **99**(3): 259-269.
- Friedrich, R., P. Fuentes-Prior, E. Ong, G. Coombs, M. Hunter, R. Oehler, D. Pierson, R. Gonzalez, R. Huber, W. Bode and E. L. Madison (2002). "Catalytic domain structures of MT-SP1/matriptase, a matrix-degrading transmembrane serine proteinase." J Biol Chem **277**(3): 2160-2168.
- Goeddel, D. V., D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura and A. D. Riggs (1979). "Expression in Escherichia coli of chemically synthesized genes for human insulin." Proc Natl Acad Sci U S A **76**(1): 106-110.
- Gou, X.-H., Y.-Y. Liu, Q.-L. Chen, J.-J. Tang, D.-Y. Liu, L. Zou, X.-Y. Wu and W. Wang (2012). "High level expression of bikunin in Pichia pastoris by fusion of human serum albumin." AMB Express **2**(1): 14.
- Granados, R. R., G. Li and G. W. Blissard (2007). "Insect cell culture and biotechnology." Virologica Sinica **22**(2): 83-93.
- Gruber, A. R., R. Lorenz, S. H. Bernhart, R. Neuböck and I. L. Hofacker (2008). "The Vienna RNA Websuite." Nucleic Acids Research **36**(Web Server issue): W70-W74.
- Guerfal, M., S. Ryckaert, P. P. Jacobs, P. Ameloot, K. Van Craenenbroeck, R. Derycke and N. Callewaert (2010). "The HAC1 gene from Pichia pastoris: characterization and effect of its overexpression on the production of secreted, surface displayed and membrane proteins." Microb Cell Fact **9**: 49.
- Hartner, F. S. and A. Glieder (2006). "Regulation of methanol utilisation pathway genes in yeasts." Microb Cell Fact **5**: 39.
- Hartner, F. S., C. Ruth, D. Langenegger, S. N. Johnson, P. Hyka, G. P. Lin-Cereghino, J. Lin-Cereghino, K. Kovar, J. M. Cregg and A. Glieder (2008). "Promoter library designed for fine-tuned gene expression in Pichia pastoris." Nucleic Acids Research **36**(12): e76.
- Hatesuer, B., S. Bertram, N. Mehnert, M. M. Bahgat, P. S. Nelson, S. Pohlman and K. Schughart (2013). "Tmprss2 is essential for influenza H1N1 virus pathogenesis in mice." PLoS Pathog **9**(12): e1003774.
- Higgins, D. R. and J. M. Cregg (1998). "Introduction to Pichia pastoris." Methods Mol Biol **103**: 1-15.
- Hitchcock, A. L., J. A. Kahana and P. A. Silver (2006). "The uses of green fluorescent protein in yeasts." Methods Biochem Anal **47**: 179-201.
- Hong, I. P., S. J. Lee, Y. S. Kim and S. G. Choi (2007). "Recombinant expression of human cathelicidin (hCAP18/LL-37) in Pichia pastoris." Biotechnol Lett **29**(1): 73-78.
- Huang, Y. S., Z. Chen, Y. Q. Chen, G. C. Ma, J. F. Shan, W. Liu and L. F. Zhou (2008). "Preparation and characterization of a novel exendin-4 human serum albumin fusion protein expressed in Pichia pastoris." J Pept Sci **14**(5): 588-595.
- Huberman, J. A., J. G. Zhu, L. R. Davis and C. S. Newlon (1988). "Close association of a DNA replication origin and an ARS element on chromosome III of the yeast, Saccharomyces cerevisiae." Nucleic Acids Res **16**(14A): 6373-6384.
- Hunt, I. (2005). "From gene to protein: a review of new and enabling technologies for multi-parallel protein expression." Protein Expr Purif **40**(1): 1-22.

- Hurt, A. C., T. Chotpitayasunondh, N. J. Cox, R. Daniels, A. M. Fry, L. V. Gubareva, F. G. Hayden, D. S. Hui, O. Hungnes, A. Lackenby, W. Lim, A. Meijer, C. Penn, M. Tashiro, T. M. Uyeki and M. Zambon (2012). "Antiviral resistance during the 2009 influenza A H1N1 pandemic: public health, laboratory, and clinical perspectives." Lancet Infect Dis **12**(3): 240-248.
- Ikonomou, L., Y. J. Schneider and S. N. Agathos (2003). "Insect cell culture for industrial production of recombinant proteins." Appl Microbiol Biotechnol **62**(1): 1-20.
- Kelley, L. A., S. Mezulis, C. M. Yates, M. N. Wass and M. J. Sternberg (2015). "The Phyre2 web portal for protein modeling, prediction and analysis." Nat Protoc **10**(6): 845-858.
- Khan, A. R. and M. N. G. James (1998). "Molecular mechanisms for the conversion of zymogens to active proteolytic enzymes." Protein Science **7**(4): 815-836.
- Kim, T. S., C. Heinlein, R. C. Hackman and P. S. Nelson (2006). "Phenotypic Analysis of Mice Lacking the Tmprss2-Encoded Protease." Molecular and Cellular Biology **26**(3): 965-975.
- Klenk, H. D. and R. Rott (1988). "The molecular biology of influenza virus pathogenicity." Adv Virus Res **34**: 247-281.
- Kontermann, R. E. (2010). "Alternative antibody formats." Curr Opin Mol Ther **12**(2): 176-183.
- Koutz, P., G. R. Davis, C. Stillman, K. Barringer, J. Cregg and G. Thill (1989). "Structural comparison of the Pichia pastoris alcohol oxidase genes." Yeast **5**(3): 167-177.
- Küberl, A., J. Schneider, G. G. Thallinger, I. Anderl, D. Wibberg, T. Hajek, S. Jaenicke, K. Brinkrolf, A. Goesmann, R. Szczepanowski, A. Puhler, H. Schwab, A. Glieder and H. Pichler (2011). "High-quality genome sequence of Pichia pastoris CBS7435." J Biotechnol **154**(4): 312-320.
- Kumar, A. and S. Singh (2013). "Directed evolution: tailoring biocatalysts for industrial applications." Crit Rev Biotechnol **33**(4): 365-378.
- Kwon, K., J. Hasseman, S. Latham, C. Grose, Y. Do, R. D. Fleischmann, R. Pieper and S. N. Peterson (2011). "Recombinant expression and functional analysis of proteases from Streptococcus pneumoniae, Bacillus anthracis, and Yersinia pestis." BMC Biochemistry **12**: 17-17.
- Kyriakopoulos, S. and C. Kontoravdi (2013). "Analysis of the landscape of biologically-derived pharmaceuticals in Europe: dominant production systems, molecule types on the rise and approval trends." Eur J Pharm Sci **48**(3): 428-441.
- Lee, C. C., T. G. Williams, D. W. S. Wong and G. H. Robertson (2005). "An episomal expression vector for screening mutant gene libraries in Pichia pastoris." Plasmid **54**(1): 80-85.
- Lee, M. S., I. C. Tseng, Y. Wang, K. Kiyomiya, M. D. Johnson, R. B. Dickson and C. Y. Lin (2007). "Autoactivation of matriptase in vitro: requirement for biomembrane and LDL receptor domain." Am J Physiol Cell Physiol **293**(1): C95-105.
- Lei, J., B. Guan, B. Li, Z. Duan, Y. Chen, H. Li and J. Jin (2012). "Expression, purification and characterization of recombinant human interleukin-2-serum albumin (rhIL-2-HSA) fusion protein in Pichia pastoris." Protein Expr Purif **84**(1): 154-160.
- Liang, S., Y. Lin, C. Li and Y. Ye (2012). "Internal ribosome entry site mediates protein synthesis in yeast Pichia pastoris." Biotechnol Lett **34**(5): 957-964.
- Liang, S., C. Zou, Y. Lin, X. Zhang and Y. Ye (2013). "Identification and characterization of P GCW14 : a novel, strong constitutive promoter of Pichia pastoris." Biotechnol Lett.
- Lin-Cereghino, G. P., J. Lin-Cereghino, A. J. Sunga, M. A. Johnson, M. Lim, M. A. Gleeson and J. M. Cregg (2001). "New selectable marker/auxotrophic host strain combinations for molecular genetic manipulation of Pichia pastoris." Gene **263**(1-2): 159-169.
- Lin-Cereghino, J., M. D. Hashimoto, A. Moy, J. Castelo, C. C. Orazem, P. Kuo, S. Xiong, V. Gandhi, C. T. Hatae, A. Chan and G. P. Lin-Cereghino (2008). "Direct selection of Pichia pastoris expression strains using new G418 resistance vectors." Yeast (Chichester, England) **25**(4): 293-299.
- Lin-Cereghino, J. and G. P. Lin-Cereghino (2007). Vectors and Strains for Expression. Pichia Protocols. J. M. Cregg. Totowa, NJ, Humana Press**:** 11-25.
- Lin-Cereghino, J., W. W. Wong, S. Xiong, W. Giang, L. T. Luong, J. Vu, S. D. Johnson and G. P. Lin-Cereghino (2005). "Condensed protocol for competent cell preparation and transformation of the methylotrophic yeast Pichia pastoris." Biotechniques **38**(1): 44, 46, 48.
- Lin, B., C. Ferguson, J. T. White, S. Wang, R. Vessella, L. D. True, L. Hood and P. S. Nelson (1999). "Prostate-localized and androgen-regulated expression of the membrane-bound serine protease TMPRSS2." Cancer Res **59**(17): 4180-4184.
- Liu, H., Y. Qin, Y. Huang, Y. Chen, P. Cong and Z. He (2014). "Direct evaluation of the effect of gene dosage on secretion of protein from yeast Pichia pastoris by expressing EGFP." J Microbiol Biotechnol **24**(2): 144-151.
- Looke, M., K. Kristjuhan and A. Kristjuhan (2011). "Extraction of genomic DNA from yeasts for PCR-based applications." Biotechniques **50**(5): 325-328.
- Ma, C.-H., A. Kwiatek, S. Bolusani, Y. Voziyanov and M. Jayaram (2007). "Unveiling hidden catalytic contributions of the conserved His/Trp-III in tyrosine recombinases: assembly of a novel active site in Flp recombinase harboring alanine at this position." Journal of molecular biology **368**(1): 183-196.
- Macauley-Patrick, S., M. L. Fazenda, B. McNeil and L. M. Harvey (2005). "Heterologous protein production using the Pichia pastoris expression system." Yeast **22**(4): 249-270.

- Maccani, A., N. Landes, G. Stadlmayr, D. Maresch, C. Leitner, M. Maurer, B. Gasser, W. Ernst, R. Kunert and D. Mattanovich (2014). "Pichia pastoris secretes recombinant proteins less efficiently than Chinese hamster ovary cells but allows higher space-time yields for less complex proteins." Biotechnol J **9**(4): 526-537.
- Makrides, S. C. (1996). "Strategies for achieving high-level expression of genes in Escherichia coli." Microbiol Rev **60**(3): 512-538.
- Meyer, M. and I. Jaspers (2015). "Respiratory protease/antiprotease balance determines susceptibility to viral infection and can be modified by nutritional antioxidants." American Journal of Physiology - Lung Cellular and Molecular Physiology **308**(12): L1189-L1201.
- Meyer, S., C. Lorenz, B. Baser, M. Wordehoff, V. Jager and J. van den Heuvel (2013). "Multi-host expression system for recombinant production of challenging proteins." PLoS One **8**(7): e68674.
- Mignone, F., C. Gissi, S. Liuni and G. Pesole (2002). "Untranslated regions of mRNAs." Genome Biology **3**(3): reviews0004.0001.
- Näätsaari, L., B. Mistlberger, C. Ruth, T. Hajek, F. S. Hartner and A. Glieder (2012). "Deletion of the *Pichia pastoris KU70* Homologue Facilitates Platform Strain Generation for Gene Expression and Synthetic Biology." PLoS ONE **7**(6): e39720.
- Nettleship, J. E., R. Assenberg, J. M. Diprose, N. Rahman-Huq and R. J. Owens (2010). "Recent advances in the production of proteins in insect and mammalian cells for structural biology." J Struct Biol **172**(1): 55-65.
- Netzel-Arnett, S., J. D. Hooper, R. Szabo, E. L. Madison, J. P. Quigley, T. H. Bugge and T. M. Antalis (2003). "Membrane anchored serine proteases: a rapidly expanding group of cell surface proteolytic enzymes with potential roles in cancer." Cancer Metastasis Rev **22**(2-3): 237-258.
- O'Gorman, S., D. T. Fox and G. M. Wahl (1991). "Recombinase-mediated gene activation and site-specific integration in mammalian cells." Science **251**(4999): 1351-1355.
- Ogata, K., H. Nishikawa and M. Ohsugi (1969). "A Yeast Capable of Utilizing Methanol." Agricultural and Biological Chemistry **33**(10): 1519-1520.
- Ormö, M., A. B. Cubitt, K. Kallio, L. A. Gross, R. Y. Tsien and S. J. Remington (1996). "Crystal structure of the Aequorea victoria green fluorescent protein." Science **273**(5280): 1392-1395.
- Papakonstantinou, T., S. Harris and M. T. W. Hearn (2009). "Expression of GFP using Pichia pastoris vectors with zeocin or G-418 sulphate as the primary selectable marker." Yeast **26**(6): 311-321.
- Papula, L. (2008). Mathematik für Ingenieure und Naturwissenschaftler, Band 3, 5. verbesserte und erweiterte Auflage, Vieweg & Teubner.
- Periyasamy, S., N. Govindappa, S. Sreenivas and K. Sastry (2013a). "Isolation, characterization and evaluation of the Pichia pastoris sorbitol dehydrogenase promoter for expression of heterologous proteins." Protein Expr Purif **92**(1): 128-133.
- Periyasamy, S., N. Govindappa, S. Sreenivas and K. Sastry (2013b). "Isolation, characterization and evaluation of the Pichia pastoris sorbitol dehydrogenase promoter for expression of heterologous proteins." Protein Expression and Purification **92**(1): 128-133.
- Pillay, K. and P. Govender (2013). "Amylin Uncovered: A Review on the Polypeptide Responsible for Type II Diabetes." BioMed Research International **2013**: 826706.
- Pla, I. A., L. M. Damasceno, T. Vannelli, G. Ritter, C. A. Batt and M. L. Shuler (2006). "Evaluation of Mut+ and MutS Pichia pastoris phenotypes for high level extracellular scFv expression under feedback control of the methanol concentration." Biotechnol Prog **22**(3): 881-888.
- Prasher, D. C., V. K. Eckenrode, W. W. Ward, F. G. Prendergast and M. J. Cormier (1992). "Primary structure of the Aequorea victoria green-fluorescent protein." Gene **111**(2): 229-233.
- Prielhofer, R., M. Maurer, J. Klein, J. Wenger, C. Kiziak, B. Gasser and D. Mattanovich (2013). "Induction without methanol: novel regulated promoters enable high-level expression in Pichia pastoris." Microb Cell Fact **12**: 5.
- Qin, X., J. Qian, G. Yao, Y. Zhuang, S. Zhang and J. Chu (2011). "GAP Promoter Library for Fine-Tuning of Gene Expression in Pichia pastoris." Applied and Environmental Microbiology **77**(11): 3600-3608.
- Ramsay, A. J., V. Quesada, M. Sanchez, C. Garabaya, M. P. Sarda, M. Baiget, A. Remacha, G. Velasco and C. Lopez-Otin (2009). "Matriptase-2 mutations in iron-refractory iron deficiency anemia patients provide new insights into protease activation mechanisms." Hum Mol Genet **18**(19): 3673-3683.
- Reitz, C. and R. Mayeux (2014). "Alzheimer disease: Epidemiology, Diagnostic Criteria, Risk Factors and Biomarkers." Biochemical pharmacology **88**(4): 640-651.
- Rhee, W. J., E. J. Kim and T. H. Park (1999). "Kinetic effect of silkworm hemolymph on the delayed host cell death in an insect cell-baculovirus system." Biotechnol Prog **15**(6): 1028-1032.
- Rodrigues, F., M. van Hemert, H. Y. Steensma, M. Corte-Real and C. Leao (2001). "Red fluorescent protein (DsRed) as a reporter in Saccharomyces cerevisiae." J Bacteriol **183**(12): 3791-3794.
- Routledge, S. J., C. J. Hewitt, N. Bora and R. M. Bill (2011). "Antifoam addition to shake flask cultures of recombinant Pichia pastoris increases yield." Microb Cell Fact **10**: 17.
- Sakai, K., Y. Ami, M. Tahara, T. Kubota, M. Anraku, M. Abe, N. Nakajima, T. Sekizuka, K. Shirato, Y. Suzaki, A. Ainai, Y. Nakatsu, K. Kanou, K. Nakamura, T. Suzuki, K. Komase, E. Nobusawa, K. Maenaka, M. Kuroda, H. Hasegawa, Y.

Kawaoka, M. Tashiro and M. Takeda (2014). "The host protease TMPRSS2 plays a major role in in vivo replication of emerging H7N9 and seasonal influenza viruses." J Virol **88**(10): 5608-5616.

- Sambrook, J., D. W. Russell (2001). Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press.
- Sambrook, J. and D. W. Russell (2006). "The inoue method for preparation and transformation of competent e. Coli: "ultra-competent" cells." CSH Protoc **2006**(1).
- Sandig, V., T. Rose, K. Winkler and R. Brecht (2005). Mammalian Cells. Production of Recombinant Proteins, Wiley-VCH Verlag GmbH & Co. KGaA**:** 233-252.
- Sauer, B. (1987). "Functional expression of the cre-lox site-specific recombination system in the yeast Saccharomyces cerevisiae." Mol Cell Biol **7**(6): 2087-2096.
- Schägger, H. and G. von Jagow (1991). "Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form." Anal Biochem **199**(2): 223-231.
- Schlake, T. and J. Bode (1994). "Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci." Biochemistry **33**(43): 12746-12751.
- Schmittgen, T. D. and K. J. Livak (2008). "Analyzing real-time PCR data by the comparative C(T) method." Nat Protoc **3**(6): 1101-1108.
- Schübeler, D., K. Maass and J. Bode (1998). "Retargeting of retroviral integration sites for the predictable expression of transgenes and the analysis of cis-acting sequences." Biochemistry **37**(34): 11907-11914.
- Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer and R. Y. Tsien (2004). "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein." Nat Biotechnol **22**(12): 1567-1572.
- Shimomura, O., F. H. Johnson and Y. Saiga (1962). "Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea." J Cell Comp Physiol **59**: 223-239.
- Shirogane, Y., M. Takeda, M. Iwasaki, N. Ishiguro, H. Takeuchi, Y. Nakatsu, M. Tahara, H. Kikuta and Y. Yanagi (2008). "Efficient multiplication of human metapneumovirus in Vero cells expressing the transmembrane serine protease TMPRSS2." J Virol **82**(17): 8942-8946.
- Silvestri, L., F. Guillem, A. Pagani, A. Nai, C. Oudin, M. Silva, F. Toutain, C. Kannengiesser, C. Beaumont, C. Camaschella and B. Grandchamp (2009). "Molecular mechanisms of the defective hepcidin inhibition in TMPRSS6 mutations associated with iron-refractory iron deficiency anemia." Blood **113**(22): 5605-5608.
- Soderholm, J., D. Bhattacharyya, D. Strongin, V. Markovitz, P. L. Connerly, C. A. Reinke and B. S. Glick (2004). "The transitional ER localization mechanism of Pichia pastoris Sec12." Dev Cell **6**(5): 649-659.
- Somoza, J. R., J. D. Ho, C. Luong, M. Ghate, P. A. Sprengeler, K. Mortara, W. D. Shrader, D. Sperandio, H. Chan, M. E. McGrath and B. A. Katz (2003). "The structure of the extracellular region of human hepsin reveals a serine protease domain and a novel scavenger receptor cysteine-rich (SRCR) domain." Structure **11**(9): 1123-1131.
- Steinhauer, D. A. (1999). "Role of hemagglutinin cleavage for the pathogenicity of influenza virus." Virology **258**(1): 1-20.
- Sturmberger, L., T. Chappell, M. Geier, F. Krainer, K. J. Day, U. Vide, S. Trstenjak, A. Schiefer, T. Richardson, L. Soriaga, B. Darnhofer, R. Birner-Gruenberger, B. S. Glick, I. Tolstorukov, J. Cregg, K. Madden and A. Glieder (2016). "Refined Pichia pastoris reference genome sequence." J Biotechnol **235**: 121-131.
- Subramani, S., A. Koller and W. B. Snyder (2000). "Import of peroxisomal matrix and membrane proteins." Annu Rev Biochem **69**: 399-418.
- Sunga, A. J., I. Tolstorukov and J. M. Cregg (2008). "Posttransformational vector amplification in the yeast Pichia pastoris." FEMS Yeast Res **8**(6): 870-876.
- Szabo, R. and T. H. Bugge (2008). "Type II transmembrane serine proteases in development and disease." Int J Biochem Cell Biol **40**(6-7): 1297-1316.
- Szabo, R., Q. Wu, R. B. Dickson, S. Netzel-Arnett, T. M. Antalis and T. H. Bugge (2003). "Type II transmembrane serine proteases." Thromb Haemost **90**(2): 185-193.
- Tarnow, C., G. Engels, A. Arendt, F. Schwalm, H. Sediri, A. Preuss, P. S. Nelson, W. Garten, H.-D. Klenk, G. Gabriel and E. Böttcher-Friebertshäuser (2014). "TMPRSS2 Is a Host Factor That Is Essential for Pneumotropism and Pathogenicity of H7N9 Influenza A Virus in Mice." Journal of Virology **88**(9): 4744-4751.
- Taubenberger, J. K. and D. M. Morens (2006). "1918 Influenza: the mother of all pandemics." Emerg Infect Dis **12**(1): 15-22.
- Trowitzsch, S., C. Bieniossek, Y. Nie, F. Garzoni and I. Berger (2010). "New baculovirus expression tools for recombinant protein complex production." J Struct Biol **172**(1): 45-54.
- Vaarala, M. H., K. S. Porvari, S. Kellokumpu, A. P. Kyllonen and P. T. Vihko (2001). "Expression of transmembrane serine protease TMPRSS2 in mouse and human tissues." J Pathol **193**(1): 134-140.
- Van Craenenbroeck, K., P. Vanhoenacker and G. Haegeman (2000). "Episomal vectors for gene expression in mammalian cells." Eur J Biochem **267**(18): 5665-5678.
- Vaughn, J. L., R. H. Goodwin, G. J. Tompkins and P. McCawley (1977). "The establishment of two cell lines from the insect Spodoptera frugiperda (Lepidoptera; Noctuidae)." In Vitro **13**(4): 213-217.

- Vedvick, T., R. G. Buckholz, M. Engel, M. Urcan, J. Kinney, S. Provow, R. S. Siegel and G. P. Thill (1991). "High-level secretion of biologically active aprotinin from the yeast Pichia pastoris." J Ind Microbiol **7**(3): 197-201.
- Verhoeyen, E., H. Hauser and D. Wirth (2001). "Evaluation of retroviral vector design in defined chromosomal loci by Flp-mediated cassette replacement." Hum Gene Ther **12**(8): 933-944.
- Walter, M., C. Chaban, K. Schutze, O. Batistic, K. Weckermann, C. Nake, D. Blazevic, C. Grefen, K. Schumacher, C. Oecking, K. Harter and J. Kudla (2004). "Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation." Plant J **40**(3): 428-438.
- Wan, H., M. Sjolinder, H. U. Schairer and A. Leclerque (2004). "A new dominant selection marker for transformation of Pichia pastoris to soraphen A resistance." J Microbiol Methods **57**(1): 33-39.
- Wegner, G. H. (1990). "Emerging applications of the methylotrophic yeasts." FEMS Microbiol Rev **7**(3-4): 279-283.
- Wen, H. and Z. Wang (2005). "Expression and characterization of rubella virus glycoprotein E1 in yeast cells." Intervirology **48**(5): 321-328.
- Weninger, A., A.-M. Hatzl, C. Schmid, T. Vogl and A. Glieder (2016). "Combinatorial optimization of CRISPR/Cas9 expression enables precision genome engineering in the methylotrophic yeast Pichia pastoris." Journal of Biotechnology **235**: 139-149.
- Wickham, T. J., T. Davis, R. R. Granados, M. L. Shuler and H. A. Wood (1992). "Screening of insect cell lines for the production of recombinant proteins and infectious virus in the baculovirus expression system." Biotechnol Prog **8**(5): 391-396.
- Wickham, T. J. and G. R. Nemerow (1993). "Optimization of growth methods and recombinant protein production in BTI-Tn-5B1-4 insect cells using the baculovirus expression system." Biotechnol Prog **9**(1): 25-30.
- Wilde, M., M. Klausberger, D. Palmberger, W. Ernst and R. Grabherr (2014). "Tnao38, high five and Sf9--evaluation of host-virus interactions in three different insect cell lines: baculovirus production and recombinant protein expression." Biotechnol Lett **36**(4): 743-749.
- Wilke, S., L. Groebe, V. Maffenbeier, V. Jäger, M. Gossen, J. Josewski, A. Duda, L. Polle, R. J. Owens, D. Wirth, D. W. Heinz, J. van den Heuvel and K. Büssow (2011). "Streamlining Homogeneous Glycoprotein Production for Biophysical and Structural Applications by Targeted Cell Line Development." PLoS ONE **6**(12): e27829.
- Wilson, S., B. Greer, J. Hooper, A. Zijlstra, B. Walker, J. Quigley and S. Hawthorne (2005). "The membrane-anchored serine protease, TMPRSS2, activates PAR-2 in prostate cancer cells." Biochem J **388**(Pt 3): 967-972.
- Wurm, F. M. (2004). "Production of recombinant protein therapeutics in cultivated mammalian cells." Nat Biotechnol **22**(11): 1393-1398.
- Yang, J., W. Jiang and S. Yang (2009). "mazF as a counter-selectable marker for unmarked genetic modification of Pichia pastoris." FEMS Yeast Research **9**(4): 600-609.
- Yang, M., S. C. Johnson and P. P. Murthy (2012). "Enhancement of alkaline phytase production in Pichia pastoris: influence of gene dosage, sequence optimization and expression temperature." Protein Expr Purif **84**(2): 247-254.
- Yang, M., S. Teymorian, P. Olivares and P. P. N. Murthy (2015). "Extracellular expression of alkaline phytase in Pichia pastoris: Influence of signal peptides, promoters and growth medium." Biotechnology Reports **6**: 112-118.
- Yasuoka, S., T. Ohnishi, S. Kawano, S. Tsuchihashi, M. Ogawara, K. Masuda, K. Yamaoka, M. Takahashi and T. Sano (1997). "Purification, characterization, and localization of a novel trypsin-like protease found in the human airway." Am J Respir Cell Mol Biol **16**(3): 300-308.
- Yuan, C., L. Chen, E. J. Meehan, N. Daly, D. J. Craik, M. Huang and J. C. Ngo (2011). "Structure of catalytic domain of Matriptase in complex with Sunflower trypsin inhibitor-1." BMC Structural Biology **11**(1): 30.
- Zhang, C. and S. H. Kim (2003). "Overview of structural genomics: from structure to function." Curr Opin Chem Biol **7**(1): 28-32.
- Zhang, X., X. Zhang, S. Liang, Y. Ye and Y. Lin (2013). "Key regulatory elements of a strong constitutive promoter, P GCW14, from Pichia pastoris." Biotechnol Lett.
- Zhao, B., C. Yuan, R. Li, D. Qu, M. Huang and J. C. Ngo (2013). "Crystal structures of matriptase in complex with its inhibitor hepatocyte growth factor activator inhibitor-1." J Biol Chem **288**(16): 11155-11164.
- Zhu, R. Y., X. Xin, H. Y. Dai, Q. Li, J. Y. Lei, Y. Chen and J. Jin (2012). "Expression and purification of recombinant human serum albumin fusion protein with VEGF165b in Pichia pastoris." Protein Expr Purif **85**(1): 32-37.

.

Appendix I: Oligonucleotides

Cloning, vector construction and mutagenesis

Name	Sequence 5′ → 3′
880NcoI-dHis4-F	TAGAGCCATGGAAGCCACCTTGTGGGATCG
BamHI-OpIE1-F	CAGTAGGATCCGCGAAACACGCACGGC
BamHI-PGK-F	CAGTAGGATCCAATTCCACGGGGTTGGG
BamHI-TEF1-F	GATCAGGATCCCCCACACACCATAGCTTC
Bsa-Flpwt-F	GACGGTCTCGAATTCAAAACGATGCCACAATTTGATATATTATGTAAAAC
BsaI-MfeI-GAP-F	GACTAGGTCTCTAATTGCAATTGGATCTTTTTTGTAGAAATGTCTTGG
BsaI-TMP(pET22B)-F	GAAGCGGTCTCGAATTCGGACAGCAACTGTTCTACGTCTGAG
BsaI-TMP-F	GAAGCGGTCTCGAATTCGGTGACAGCAACTGTTCTACGTCTGAG
BstBI-HSA-FOR	CATAGTTCGAAACGATGAAGTGGGTAACCTTTATTTCC
BstBI-PGAP-ATG-R	CATTCTTCGAACCATATGCCATATAGTTGTTCAATTGATTGAAATAGG
Cyc1T-NdeI-R	GATCGCATATGAGCTTGCAAATTAAAGCCTTCGAG
dHis4-880NcoI-R	CAGTCCCATGGCTCTAAGACCCTTTG
dHis4T-ScaI-R	GCTCAAGTACTTAATGCGGTAGTTTATCACAGTTAAATTGCTAACGC
EcoRI-GAP-R	GTCTAGAATTCATAGTTGTTCAATTCATTGAAATAGG
EcoRI-mCherry-f	GCTACGAATTCAGCATGGTGAGCAAGGGCGAGGAGGATAAC
EcoRI-Tmprss2-opt-F	GAAGCTGAATTCGGTGACTCAAATTGCTC
EcoRI-yeGFP-F	GCAGGAATTCATGTCTAAAGGTGAAGAATTATTCACC
GAPintF3-F	GAAGTTCCTATTCCGAAGTTCCTATTCTTCAAATAGTATAGGAACTTCAGAATATAAAAGGCGAACACCTTTCC
GAPintF3-R	TGAAGTTCCTATACTATTTGAAGAATAGGAACTTCGGAATAGGAACTTCGATTCTGGTGGTTTCCAATAATCTC
HindIII-dHis4-F	GACTGAAGCTTCTGGTTCGGGAACATTTCCCTTGCTACCTGC
int-TMPopt-mut-F	GTAAGACTAAAAACAACGACATCGCTTTGATGAAGCTTCAGAC
int-TMPopt-mut-R	CTTCATCAAAGCGATGTCGTTGTTTTTAGTCTTACTATCGTAG
kan_r1	GTCGGGCCGCGTCGGACGTGCTGCAGTTAGAAAAACTCATCGAGCATC
NcoI-kan_f	GCAGACCATGGATGCTGATTTATATGGGTATAAATGGGC
NheI-TMP-F	GATAGGCTAGCTGGTGACAGCAACTGTTCTACGTCTGAG
Not-Flpwt-R	CATGCGGCCGCTTAACCCGCGGATATGCGTCTATTTATGTAGGATGAAAG
NotI-mCh(woTAG)-r	GTTATGCGGCCGCCTTGTACAGCTCGTCCATGCCG
NotI-Rudolph-R	CATGAGCGGCCGCAGTTTCCTTAACATCTACTGTGAAGTG
NotI-TMP-R2	CTACTTGCGGCCGCTCAGTGATGGTGATGGTGATGGTGATGCTTATCATCATCATCCTTGTAATCGC
NotI-Tmprss2-opt-R	CTAGATGCGGCCGCCTAATGGTG
Not-Spe-F3_R	GTATTGCGGCCGCATAGTTCAATACTAGTGAAGTTCCTATACTATTTGAAGAATAG
PciI-3'AOX-f	GCATTACATGTGTCGGCATCACCGGCG
PciI-3'AOX-r	CAATCACATGTTCTTTCCTGCGTTATCCCCTG
PGK-PstI-R	GATCTCTGCAGCCTGGGGAGAGAGGTCG
PstI-OpIE1-R	GATCTCTGCAGGCAAAGGTGCTGCGCG
SacII-yeGFP-R	CTACACCGCGGTTTGTACAAAAGCTCCATACCATGAG
ScaI-kan_r2	CTCTCAGTACTCGGACCCGTCGGGCCGCGTCGGACGTG
SpeI-PGAP-F	GATACACTAGTAGATCTTTTTTGTAGAAATGTCTTGG
SpeI-URG-F3-R2	GATCTACTAGTGAAGTTCCTATACTATTTGAAGAATAGGAACTTCGGAATAGGAAC
TMP-R1	CTTATCATCATCATCCTTGTAATCGCATGCACCTTGAAAGTACAAGTTTTCGCTGTTCGCCCTCATTTGC
URG-F3-R1	GAATAGGAACTTCGGAATAGGAACTTCCTCGAGTTGTCAAGTGAAAGGGGCAATT
XbaI-URG-F	CACTATCTAGACTGCAGCTGCTACTCTGGTCCCAAGTGAAC
XhoI-HSA-REV	GTATACTCGAGTGCATGCACGTGAATTCTGGACCTTGAAACAAAACTTCCAATAAGCCTAAGGCAGCTTGAC

RT-PCR

Name	Sequence 5′ → 3′
rt_Rud-F	GTGATGGTATTCTCAAAGGCGACGTTAC
rt_Rud-R	GGACATGCTCGATAATGTGATTTGGTGG
rt_PGK1-F	TAAGGAGTTGACATACTTCGCTAAGGCC
rt_PGK1-R	CAATGATGATGGAATCGACCTTGTCCAG
rt_ACT1-F	CACAGTGTTCCCATCGGTCGTAGGTAG
rt_ACT1-R	GATACCGTGCTCGATTGGGTATCTCAAG

Sequencing and genomic control PCR

Name	Sequence 5′ → 3′
s_pPICZ-F	GACTGGTTCCAATTGACAAGC
s_pPICZ-R	GCAAATGGCATTCTGACATCC
s_pGAPZ-F	GTCCCTATTTCAATCAATTGAA
s_dHis4F1	ACATTTCCCTTGCTACCTGCATACGC
s_dHis4F2	GGAGCTGATCGACAACTTGC
s_dHis4F3	GGAAGCTGATGAACTTGCAGAAGC
s_dHis4F4	GCTGCTATGGCTTACGGAACAG
s_dHis4F5	GGTTCGTATGGAGAAACTGGGAC
s_pTEF_f	GCAATCTAATCTAAGGGCGGTG
s_tCYC1_r	GGCGTGAATGTAAGCGTGAC
s_aMF-FOR	GAAGGGGATTTCGATGTTGC
s_mCher5'-REV	CCTCCATGTGCACCTTGAAGC
s_pBRori-FOR	CGTCTTGAGTCCAACCCGGTAAG
s_AOX1TT-F	GCAGCTGATGAATATCTTGTGGTAGG
s_AmpR-F	CTCATGAGCGGATACATATTTGAATG
s_pUC-F	GGATCTCAAGAAGATCCTTTGATC
s_3'AOX1-F	CCAACATGTGTCGGCATCACC
s_mCher5'-R	CCTCCATGTGCACCTTGAAGC
s_pBRori-REV	GCCTATGGAAAAACGCCAGCAAC
s_pJET-FOR	CGACTCACTATAGGGAGAGCGGC
s_pJET-REV	AAGAACATCGATTTTCCATGGCAG
s_his4_5'-R	CAAGGAATTATCAAACTCTGCAACAC
s_gGAP-F	CGACTGTCAATCATTCATCCTTGC

Appendix II: Genomic PCR analysis

Overview of the expected band sizes of the genomic PCR analyses

The analysis of the *P. pastoris* RMCE master cell lines and the fluorochrome test cell lines was conducted in two PCR reactions. The first PCR (PCR1) was conducted to validate the integration of the vector into *P. pastoris*. The second PCR (PCR2) was conducted in parallel to distinguish single copy insertion clones from multi copy clones. The table columns indicate the vector stably transformed into *P. pastoris* by homologous recombination, the PCR oligonucleotide primers and the expected product size in base pairs [bp].

Transformed vector	PCR1		PCR2	
	Primers	Product size [bp]	Primers	Product size [bp]
pYTA-RudolphRFP	s_pGAPZ-F s_pPICZ-R	958	s_AmpR-F s_pPICZ-R	1,503
pYEXs-yeGFP	s_AOX1TT-F s_his4_5'-R	757	s_AmpR-F s_pPICZ-R	1,562
pYEXsPGK-yeGFP	s_AOX1TT-F s_his4_5'-R	783	s_AmpR-F s_pPICZ-R	1,562
pYEXsOpi-yeGFP	s_AOX1TT-F s_his4_5'-R	555	s_AmpR-F s_pPICZ-R	1,562
pPICZ-mCherry	s_pPICZ-F s_pPICZ-R	998	s_pPUC-F s_pPICZ-R	1,941
pPICZ-yeGFP	s_pPICZ-F s_pPICZ-R	1,004	s_pPUC-F s_pPICZ-R	1,947
pPICZ-RudolphRFP	s_pPICZ-F s_pPICZ-R	968	s_pPUC-F s_pPICZ-R	1,911
pGAPZAΔ8-mCherry	s_pGAPZ-F s_pPICZ-R	941	s_AmpR-F s_pPICZ-R	1,488
pGAPZAΔ8-yeGFP	s_pGAPZ-F s_pPICZ-R	947	s_AmpR-F s_pPICZ-R	1,494
pGAPZAΔ8-RudolphRFP	s_pGAPZ-F s_pPICZ-R	914	s_AmpR-F s_pPICZ-R	1,458
pYTAaox-mCherry	s_pPICZ-F s_pPICZ-R	1,015	s_3'AOX1-F s_mCher5'-R	2,077
pYTAaox-GAP-mCherry AOX1 locus insertion	s_pGAPZ-F s_pPICZ-R	941	s_3'AOX1-F s_mCher5'-R	2,570
pYTAaox-GAP-mCherry GAP locus insertion	s_pGAPZ-F s_pPICZ-R	941	s_AmpR-F s_pPICZ-R	2,478
pYTA-RudolphRFP (V1) (Repeated analysis)	s_gGAP-F s_pPICZ-R	2,063	s_AmpR-F s_pPICZ-R	1,497
pYTAΔF3-RudolphRFP (V2)	s_gGAP-F s_pPICZ-R	2,014	s_AmpR-F s_pPICZ-R	1,448
pYTAF3PR-RudolphRFP (V3)	s_gGAP-F s_pPICZ-R	2,063	s_AmpR-F s_pPICZ-R	1,497
pYTAatgF3-RudolphRFP (V4)	s_gGAP-F s_pPICZ-R	2,074	s_AmpR-F s_pPICZ-R	1,508
pYTAUR-RudolphRFP (V5)	s_gGAP-F s_pPICZ-R	2,074	s_AmpR-F s_pPICZ-R	2,022
pGAPZAΔ8-RudolphRFP (PC) (Repeated analysis)	s_gGAP-F s_pPICZ-R	2,014	s_AmpR-F s_pPICZ-R	1,448
pYTA (empty) (NC)	s_gGAP-F s_pPICZ-R	1,440	s_AmpR-F s_pPICZ-R	874

Appendix III: Cell culture data (Tmprss2)

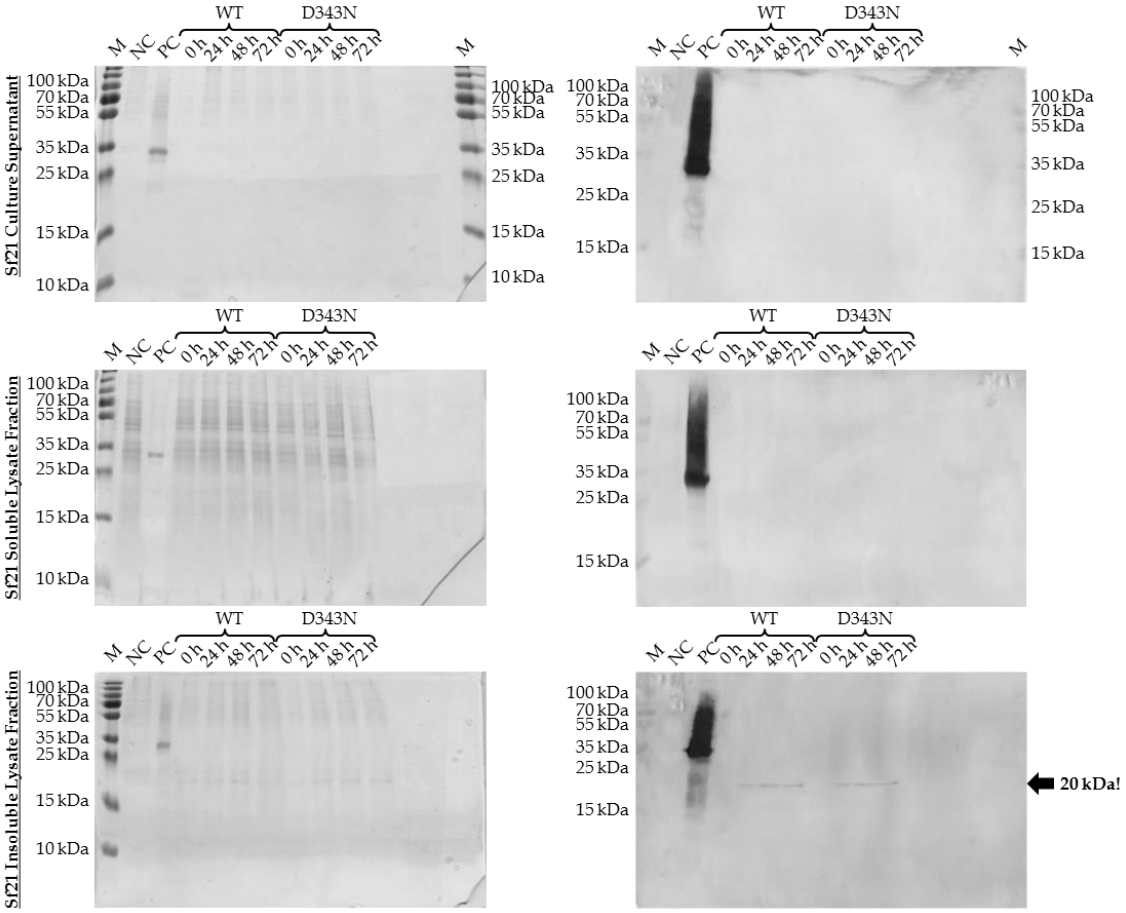
1. Growth parameters of HEK293-6E cells following transfection to express Tmprss2

HEK293-6E						
hpt	GFP gated [%]		Vitality [%]		Cells (x 10^6)	
	WT	D343N	WT	D343N	WT	D343N
0	0,02	0,02	96,8	96,8	1,6	1,6
24	-	-	-	-	2,1	2,12
48	60,73	65,82	94,1	93,7	14,6	14,8
72	78,15	81,27	92,2	90,9	15,1	15,7
96	91,22	92,65	81,5	89,7	15,1	15,3

2. Growth parameters of SF21/Hi5 cells following the infection to express Tmprss2

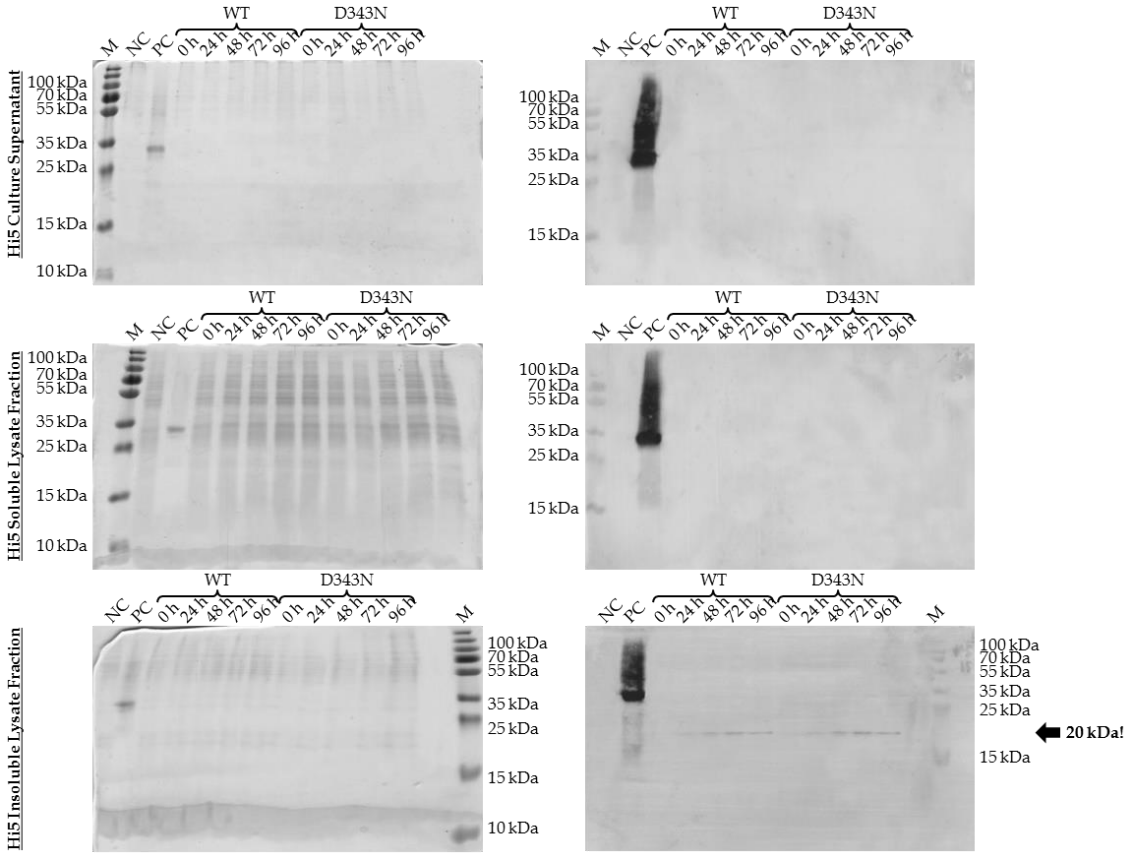
SF21						
Hours post infection	YFP gated [%]		Mean Diameter [µm]		Peak Diameter [µm]	
	WT	D343N	WT	D343N	WT	D343N
0	0,41	0,41	18,24	18,24	17,19	17,19
24	88,28	69,73	20,34	20,24	19,74	19,71
48	98,28	93,18	21,22	21,41	20,82	21,02
72	97,66	89,42	21,52	21,91	21,02	21,51
Hours post infection	Cell Count (x 10^6/mL)		Vitality [%]			
	WT	D343N	WT	D343N		
0	0,91	0,91	97	97		
24	1,3	1,2	97,1	96,8		
48	1,3	1,3	96,4	95,6		
72	1,33	1,35	96,1	94,3		
Hi5						
Hours post infection	YFP gated [%]		Mean Diameter [µm]		Peak Diameter [µm]	
	WT	D343N	WT	D343N	WT	D343N
0	0,75	0,8	24,98	24,98	24	24
24	41,69	37,44	26,45	26,41	25,44	25,54
48	61,97	51,92	27,43	27,11	26,01	26,72
72	61,97	51,92	27,81	27,67	26,46	26,97
96	44,6	43,7	27,98	27,87	26,62	27,01
Hours post infection	Cell Count (x 10^6/mL)		Vitality [%]			
	WT	D343N	WT	D343N		
0	1,1	1,1	97	97		
24	1,3	1,4	96,7	97		
48	1,4	1,45	96,7	96,5		
72	1,42	1,46	95,5	95,3		
96	1,49	1,47	81,5	89,7	15,1	15,3

3. SDS-PAGE and Western blot of the Tmprss2 expression test in SF21 cells



The unconcentrated and the TCA-precipitated supernatant samples as well as the soluble and insoluble cell lysate fractions were analyzed for the presence of Tmprss2 by SDS-PAGE and Western blot. 0 – 72 h indicate probing times in hours post infection. Neither Tmprss2-D343N nor Tmprss2-WT could be detected. Staining of the SDS gels was performed with InstantBlue. The Western blots were performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Recombnant Tmprss2-D343N from *P. pastoris* KM71H/pPICZα-Tmprss2-D343N clone 116 (processed) served as a positive control (PC). Mw standard (M): PageRuler Prestained Plus.

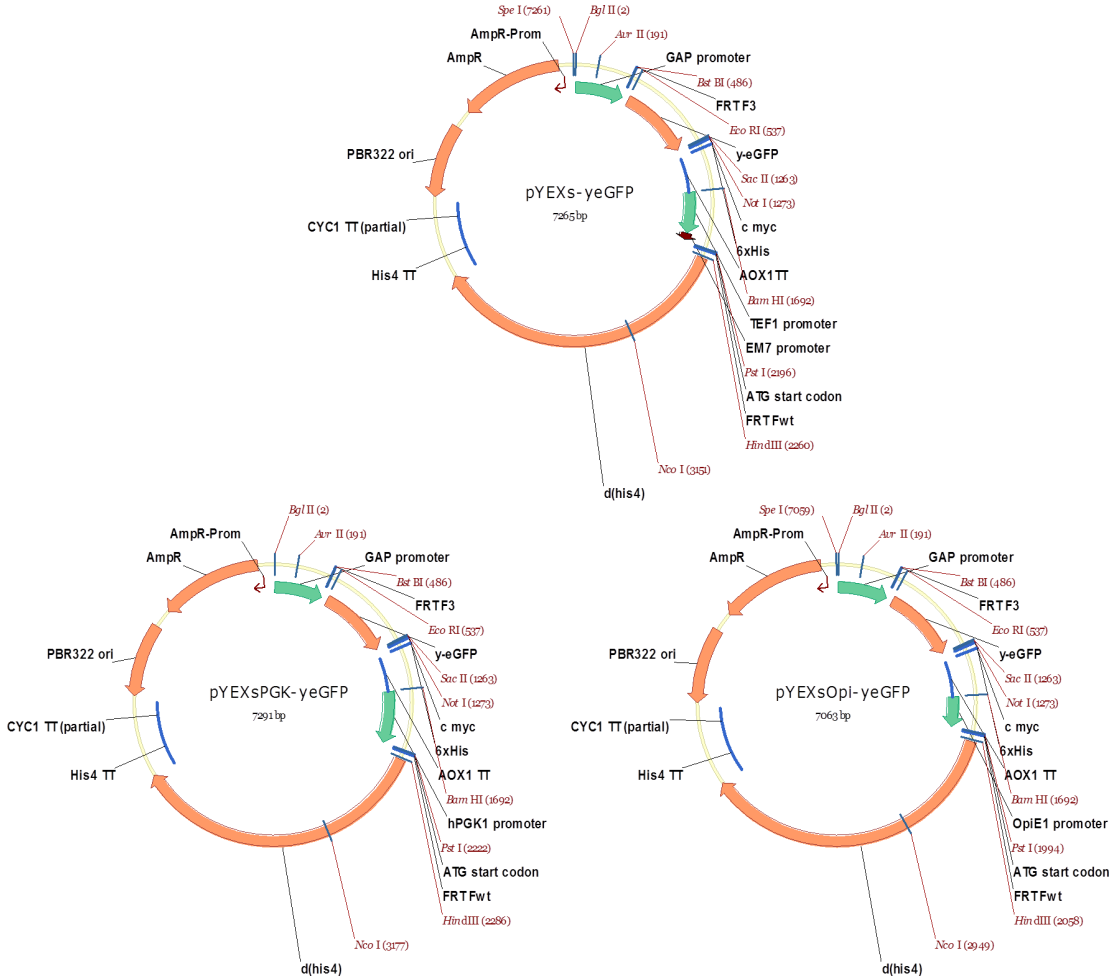
4. SDS-PAGE and Western blot of the Tmprss2 expression test in Hi5 cells



The unconcentrated and the TCA-precipitated supernatant samples as well as the soluble and insoluble cell lysate fractions were analyzed for the presence of Tmprss2 by SDS-PAGE and Western blot. 0 –96 h indicate probing times in hours post infection. Neither Tmprss2-D343N nor Tmprss2-WT could be detected. Staining of the SDS gels was performed with InstantBlue. The Western blots were performed with a primary mouse α-Flag mAb and a secondary goat AP-conjugated α-mouse pAb. Recombnant Tmprss2-D343N from *P. pastoris* KM71H/pPICZα-Tmprss2-D343N clone 116 (processed) served as a positive control (PC). Mw standard (M): PageRuler Prestained Plus.

Appendix IV: Maps of the RMCE vectors

1. pYEXs vectors (§ 3.1.3)

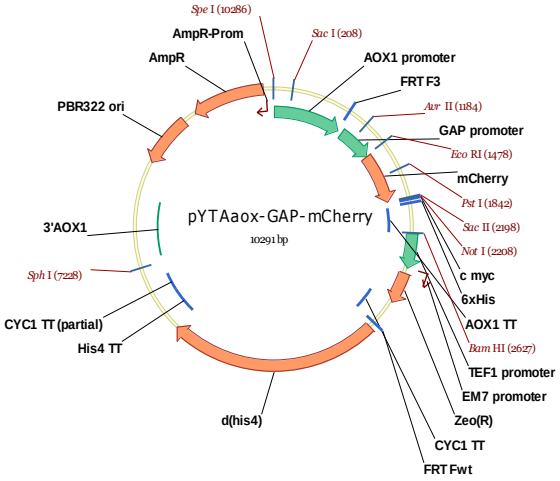


The RMCE "exchange simulation" vector pYEXs-yeGFP harbors the TEF1/EM7 promoter for the complementing expression of Δhis4 (d(his4)). In pYEXsPGK-yeGFP and pYEXsOpI-yeGFP, the TEF1/EM7 promoter was replaced by the human PGK1 (hPGK1) promoter or the OpIE1 promoter, respectively. The detailed cloning strategy is described in chapter 2.3.

2. Fluorescence test vectors (§ 3.1.4.1)

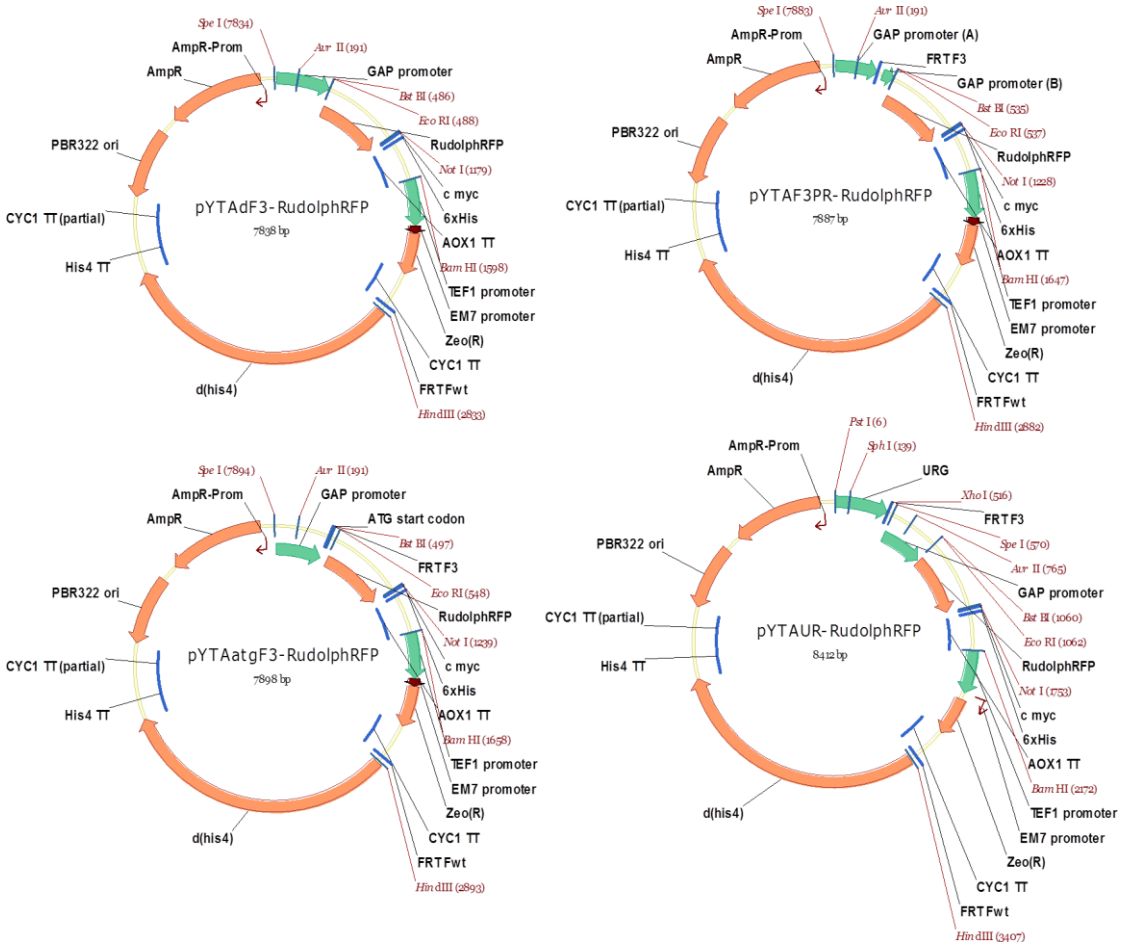
In the top row, the AOX1 RMCE tagging vector pYTAaox-mCherry is shown. Below, the respective test vectors are depicted based on pPICZ and pGAPZAΔ8. The initial tagging vector pYTA-RudolphRFP is shown for comparison. The vector maps were generated with the software Vector NTI Suite 8 (Thermo Fisher Scientific). The "d" in the figures stands for Δ. The cloning strategy is described in chapter 2.3.

3. pYTAaox-GAP-mCherry (§ 3.1.4.2)



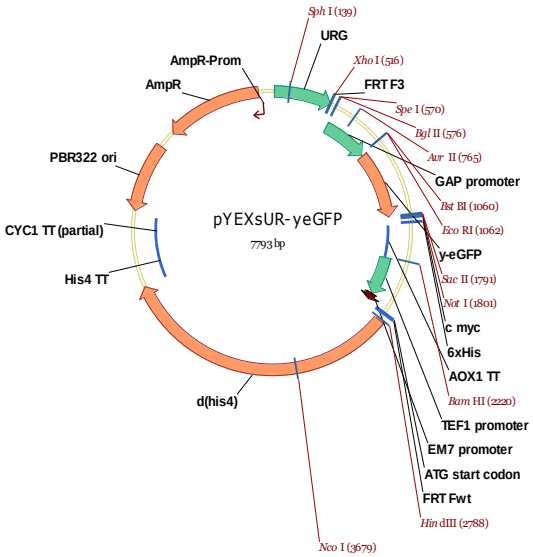
The RMCE tagging vector pYTAaox-GAP-mCherry harbors the *GAP* promoter downstream of the *AOX1* promoter and FRT F_3. The vector map was generated with Vector NTI Suite 8 (Thermo Fisher Scientific). The "d" in the figures stands for Δ. The detailed cloning strategy is described in chapter 2.3.

4. RMCE tagging vector variants for the expression analysis (§ 3.1.4.3)



The additional RMCE tagging vector variants V2 to V5 are shown: V2: pYTAΔF3-RudolphRFP; V3: pYTAF3PR-RudolphRFP; V4: pYTAatgF3-RudolphRFP; V5: pYTAUR-RudolphRFP. The "empty" pYTA was used as the negative control vector. The vector maps were generated with Vector NTI Suite 8 (Thermo Fisher Scientific). The "d" in the figures stands for Δ. The detailed cloning strategy is described in chapter 2.3.

5. pYEXsUR-yeGFP (§ 3.1.5)



The RMCE "exchange simulation"vector pYEXsUR-yeGFP harbors the TEF1/EM7 promoter for the complementing expression of Δhis4 (d(his4)) and the URG (upstream region of *GAP*) for the insertion of the vector into the genomic *GAP* locus of *P. pastoris* by homologous recombination. The vector map was generated with Vector NTI Suite 8 (Thermo Fisher Scientific). The "d" in the figures stands for Δ. The detailed cloning strategy is described in chapter 2.3.

Danksagung

Zunächst möchte ich Prof. Dr. Ralf-Rainer Mendel meinen herzlichen Dank für die Betreuung meiner Promotion als Mentor aussprechen, die ich sehr zu schätzen weiß. Ebenso möchte ich mich sehr bei Prof. Dr. Wulf Blankenfeldt und Prof. Dr. Michael Hust für die Übernahme des Korreferates und des Prüfungsvorsitzes bedanken.

Ganz besonders möchte ich mich bei Dr. Joop van den Heuvel bedanken, unter dessen Betreuung diese Arbeit am Helmholtz-Zentrum für Infektionsforschung in Braunschweig durchgeführt wurde. Er hat mir nicht nur die Gelegenheit gegeben in seiner Arbeitsgruppe mit diesen höchst interessanten und herausfordernden Projekten zu promovieren, sondern meine Arbeit auch exzellent betreut und unterstützt. Er zeigte immer großes Interesse für die Projekte und war für Fragen und Diskussionen im Rahmen dieser Doktorarbeit und darüber hinaus stets erreichbar.

Darüber hinaus möchte ich mich bei allen aktuellen und ehemaligen Mitarbeitern der Arbeitsgruppe RPEX, mit denen ich die Freude hatte zusammenzuarbeiten, sehr bedanken. Insbesondere Dr. Johannes Spehr, Dr. Bahar Baser und Dr. Maren Bleckmann, Margitta Schürig und Katharina Karste danke ich sehr für die zahlreichen wissenschaftlichen Diskussionen, Tipps und Unterstützungen. Weiterhin gilt mein ganz besonderer Dank Daniela Gebauer, Nadine Konisch und Anke Samuels für die stets exzellente technische Unterstützung im Labor, die interessanten Gespräche rund um die Arbeit und das Leben sowie die immer benötigte Portion Humor, um selbst die sporadischen grauen Labortage gut gelaunt zu überstehen. Sascha Wani, den ich während seiner Bachelorarbeit betreuen durfte, möchte ich ebenfalls meinen Dank für seine ausgezeichnete Arbeit im Labor aussprechen. Yasmin Wenzel und Vanessa Stiller, die als Azubis sehr tatkräftig die von mir bearbeiteten Projekte unterstützt haben, möchte ich in diesem Zuge auch sehr danken.

Ein großes Dankeschön geht an ebenfalls an alle Mitarbeiterinnen und Mitarbeiter der Arbeitsgruppen NBSC, SBAU und SFPR für das großartige Arbeitsklima und die sehr gute Zusammenarbeit.

Für ihre Unterstützung im Tmprss2-Projekt möchte ich des Weiteren Dr. Ruth Lambertz und Dr. Jörn Krauße danken. Darüber hinaus möchte ich Dr. Domenica Hamisch für die hilfreichen Tipps zum Thema RT-PCR danken.

Mein herzlichster Dank geht an Leonie Dalisdas für die wundervolle Zeit zwischen der Arbeit, das geduldige Zuhören und den großen Beistand in all den schwierigen Phasen. Mein größter Dank gilt meiner Familie, meinen Eltern und meiner Schwester, die mich auf diesem Weg stets unterstützt und mir diese lange Ausbildung erst ermöglicht haben. Sie tragen daher einen wesentlichen Anteil am Erfolg dieser Arbeit. Ich danke euch sehr!